APPROACHING EXACT QUANTUM CHEMISTRY BY STOCHASTIC WAVE
FUNCTION SAMPLING AND DETERMINISTIC COUPLED-CLUSTER
COMPUTATIONS

By

Jorge Emiliano Deustua Stahr

A DISSERTATION

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

Chemistry – Doctor of Philosophy

2020

ABSTRACT

APPROACHING EXACT QUANTUM CHEMISTRY BY STOCHASTIC WAVE FUNCTION SAMPLING AND DETERMINISTIC COUPLED-CLUSTER COMPUTATIONS

By

Jorge Emiliano Deustua Stahr

One of the main goals of quantum chemistry is the accurate description of ground- and excited-state energetics of increasingly complex polyatomic systems, especially when non-equilibrium structures and systems with stronger electron correlations are examined. Size extensive methods based on coupled-cluster (CC) theory and their extensions to excited electronic states via the equation-of-motion (EOM) framework have become the *de facto* standards for addressing this goal. In the vast majority of chemistry problems, the traditional single-reference CC hierarchy, including CCSD, CCSDT, CCSDTQ, etc., and its EOM counterpart provide the fastest convergence toward the exact, full configuration interaction (FCI), limit, allowing one to capture the leading many-electron correlation effects, in a systematic manner, by employing particle-hole excitations from a single reference determinant. Unfortunately, computational costs associated with the incorporation of higher-than-two-body components of the cluster and excitation operators of CC and EOMCC, which are required to achieve a fully quantitative description of molecular systems, are usually prohibitive. This has exacerbated the need for new ideas in quantum chemistry leading to computationally affordable methodologies that do not suffer from failures of lower-order, less expensive approximations. In this dissertation, I propose two novel and computationally cost-effective strategies for acquiring accurate electronic energetics, equivalent to those obtained with high-level CC/EOMCC methods, such as CCSDT, EOMCCSDT, and CCSDTQ, and the exact, FCI, theory, even when wave function quasi-degeneracies and other higher-order correlation effects become significant.

The first strategy consists in merging the deterministic CC framework, abbreviated as, CC(P;Q), with the stochastic CI Quantum Monte Carlo (QMC) approaches, such as CISDT-MC, CISDTQ-MC, FCIQMC, and their CCSDT-MC counterpart with the intention of using wave functions resulting from QMC propagations to identify the leading higher-than-doubly excited components entering deterministic CC and EOMCC calculations.

The second proposed methodology provides a direct way of recovering FCI energetics following the ideas of the externally corrected CC theories. In the resulting approach, that we denominate CAD-FCIQMC, we utilize the cluster analysis of FCIQMC wave functions to obtain the connected three- and four-body cluster components, which are the only clusters that directly couple to the one- and two-body components of the cluster operator. In general, the externally corrected CC methods are guaranteed to attain the exact, FCI, ground-state energy, if the triply and quadruply excited cluster components extracted from a non-CC source entering the suitably corrected CCSD-like equations approach their exact values. Since the long-time FCIQMC propagations converge the exact FCI wave functions, our new CAD-FCIQMC method is assured to produce numerically exact energies.

All of the above approaches, including the semi-stochastic CC(P;Q) and CAD-FCIQMC methodologies, are discussed in this dissertation. This includes their mathematical foundations, computer implementation, and numerical tests using bond breaking in the F_2, H_2O, and CH^+ molecules and the automerization of cyclobutadiene.

In memory of My Aunt, Monica Klien.

ACKNOWLEDGEMENTS

I am deeply grateful for the opportunity I have had in Professor Piecuch group in meeting and befriending my fellow group members: Dr. Adeayo O. Ajala, Dr. Nicholas P. Bauman, Mr. Arnab Chakraborty, Mr. Karthik Gururangan, Dr. Jared A. Hansen, Mr. Ilias Magoulas, Dr. Jun Shen, and Mr. Stephen H. Yuwono. My experience in the Ph.D. program would not have been the same without their support, help, encouragement, and amazing discussions.

Furthermore, I would like to thank Dr. Jonathan Moussa, who has been my MolSSI mentor for the past two years, and has helped me to a great extend in my software development efforts. Likewise, I would like to express my gratitude toward Professor Katharine C. Hunt for her encouraging attitude toward teaching and scholarship. Being able to interact with her as a scientist, Ph.D. student, and teaching assistant, has been a truly enriching and pleasurable experience.

Finally, I would like to thank my family and friends for the continuous support throughout my career, their kindness, and their love. Having such a wonderful group of people supporting me is invaluable and deeply moving.

Thank you all!

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The non-relativistic Schrödinger equation [2–7] and its relativistic extensions [8–11] are central equations in many areas of the natural sciences. This is especially true in theoretical chemistry due to the fact that they provide an accurate microscopic description of the observable properties characterizing chemical systems. In fact, being able to solve the electronic Schrödinger equation for a system consisting of any number and arrangement of atoms would allow the scientific community to greatly advance the fields of molecular and condensed matter science and, indeed, finding a way to do so is the primary goal of quantum chemistry (QC). Unfortunately, solving the electronic Schrödinger equation exactly using purely analytical means is only possible for one-electron systems, such as the hydrogen atom [3] or the H_2^+ molecule [12]. Thus, in order to study many-electron systems, which compose the vast majority of chemically interesting problems, one has to think of solving the Schrödinger equation numerically using suitably designed approximations and computer simulations. Nonetheless, achieving such objective without incurring unfeasible computational costs remains extremely difficult, which motivates the need for the development of increasingly accurate and tractable QC methods and algorithms.

Many interesting approaches exist in QC to determine, or at least approximate, the solutions of the many-electron Schrödinger equation, but two of them are of particular importance. The first one is the full configuration interaction (FCI) method, which refers to the exact diagonalization of the Hamiltonian matrix in a N-electron Hilbert space. Although FCI provides exact numerical solutions within a basis set constructed from single-particle functions, i.e., spin-orbitals, its computational cost is usually prohibitively expensive. Its applicability to systems larger than a few electrons described in more sizable and realistic sets of basis functions is beyond the reach of the most pow-

erful supercomputers in the world, which renders this method unusable for most of the present-day chemistry problems. More importantly, despite the fact that computer technology, including hardware and software, keeps improving all the time, it is extremely unlikely that it alone will outpace the steep factorial scaling associated with the dimensionalities of the FCI diagonalization problem. This issue is further exacerbated by the realization that the so-called Moore's law, introduced in 1965, predicting two-fold CPU performance gain every 18 months, may be hitting a thermodynamic limit. This hindrance forces the acquisition of more computer power away from sequential computations on single cores toward the exploitation of parallel systems and algorithms, which remain scalable, but are oftentimes quite difficult to implement, especially when coding and executing highly complex problems of the FCI type across many multi-core nodes. The promise of quantum computers that could handle FCI exists, but it is still only a promise. Even if we assume major technological advances, it is still worth rethinking the FCI approach, since much of its high cost originates from the enormous dimensionality of the many-electron Hilbert space without taking advantage of the sparsity of the Hamiltonians used in QC, which contain only pairwise electron-electron interactions. The approaches developed in this thesis, in addition to addressing issues of dimensionality, utilize the Hamiltonian's sparsity at the same time.

On the other end of the QC spectrum, the simplest and computationally least expensive approximation is the Hartree–Fock (HF) approach. The HF approach applies the variational principle to a single Slater determinant via the so-called self-consistent field (SCF) procedure, which results, at least in principle, in the energetically best zeroth-order mean-field approximation to the ground-state electronic wave function. Unfortunately, the HF method is incapable of describing many-electron correlation effects and, thus, is unable to provide vital information relevant to most chemical problems, particularly when studying chemical reactivity, weak noncovalent interactions, and even photochemistry. As an example, HF fails to capture the binding in the F_2 molecule, pro-

2

ducing a qualitatively wrong description. The restricted HF (RHF) approach overbinds the molecule, trying to dissociate it into ions, while unrestricted HF (UHF) calculations do not bind the system at all. Also, according to the HF description, the helium, beryllium, and magnesium dimers should not exist, and yet they do. Despite these and many other negative characteristics, the HF determinant serves as a convenient zeroth-order state upon which more sophisticated wave-function-based many-body methods are formulated, and is, therefore, widely used as the *de facto* standard many-electron reference function.

As already alluded to above, the HF and FCI approaches define the two extremes when approximating solutions to the Schrödinger equation, numerically, within a basis set. HF corresponds to the smallest possible CI vector, consisting of only one determinant, while the FCI vector spans the complete space of all Slater determinants. HF is inexpensive, but also inaccurate, since it ignores many-electron correlation effects. FCI describes them exactly in a given basis, but it is prohibitively expensive. We need methods in-between, which are very close to FCI at the faction of the cost.

Within the large diversity of methods one can contemplate, in which representative examples are the truncated CI hierarchy, approaches resulting from the many-body perturbation theory (MBPT), and the density-matrix renormalization group methodology [13–18], to name a few, two treatments have received most attention, especially in the recent three decades. One of them is the density functional theory (DFT) [19, 20], which is used extensively in the scientific literature mainly due to its wide availability for molecules and condensed matter systems and relatively low computational cost similar to HF calculations [21, 22]. The other one is known as the coupled-cluster (CC) theory [23–27], which is considered the best candidate for the development of methodologies that accurately approximate solutions to the Schrödinger equation, because it provides a rigorous *ab initio* framework that enables one to achieve the best balance between accuracy and computer cost in a systematically improvable and size extensive manner.

DFT approaches can achieve accurate results at a very reasonable computational cost. The scalability of DFT calculations, which allows for the routine treatment of large molecules and condensed-matter systems [21, 22, 28], enables studies in QC which are not feasible using higher-level *ab initio* wave function methodologies. Nonetheless, DFT methods usually fail in more challenging situations, such as covalent bond breaking, band gaps of bulk solids, van der Waals interactions, strongly correlated systems, and excited electronic states having double excitation or charge-transfer character [21]. In addition to this, DFT methods rely on an unknown exchange-correlation potential for which numerous approximations exist that usually require fitting to empirical or *ab initio* reference data [29], which may not be relevant to the problem of interest, putting the transferability of DFT approaches in question, especially when a diverse set of systems and phenomena needs to be explored [21, 29]. In summary, DFT remains to be a largely semi-empirical theory that provides no simple and systematic routes for improvement, in spite of its wide use and lower computational cost.

The alternative to DFT, which allows one to obtain accurate information in a systematically improvable manner and which is characterized by the fastest convergence toward the exact FCI limit, while maintaining reasonable computer effort requirements, is the CC theory. Formally, the CC theory is obtained as a generalization of MBPT by summing linked wave function and connected energy diagrams to infinite order, with the help of the linked [30–33] and connected [32, 33] cluster theorems, resulting in the iconic exponential wave function ansatz, $|\Psi_0\rangle = e^T |\Phi\rangle$ [24, 25, 32, 33]. Putting aside the basis set issue, the reliability of CC theory depends only on the quality of the reference wave function $|\Phi\rangle$, in the majority of applications the HF determinant, and its level of approximation, which is usually accomplished by truncating the cluster operator T at a given excitation rank. When T is truncated at the two-body component T_2, the resulting approximation is called CCSD (CC theory with singles and doubles) [34, 35], when it is truncated at the three-body components T_3, we obtain CCSDT (CC theory with singles,

doubles, and triples) [36, 37], when it is truncated at the four-body component T_4, the approach that we obtain is called CCSDTQ (CC theory with singles, doubles, triples, and quadruples) [38–41], etc. When the cluster operator T includes all many-body components T_n, with $n = 1, \ldots, N$, where N is the number of electrons in the system, we obtain full CC, which is equivalent to FCI. The CC approaches satisfy several important conditions of the exact theory, such as size-extensivity and wave function separability (the latter when the reference stater $|\Phi\rangle$ separates correctly). The former condition guarantees that no loss of accuracy occurs when the quantum-mechanical system of interest becomes larger; the latter that fragmentation phenomena are correctly described when the system is separated into the noninteracting subsystems. These facts are particularly important when comparing CC with truncated CI and finite-order MBPT methods (see Figure 1.1), because truncated CI is neither size extensive nor separable, whereas finite-order MBPT, which is size extensive, fails to properly describe wave function fragmentation in a non interacting subsystem limit. What is most important from a purely pragmatic point of view, the CC approximations are characterized by the very fast convergence toward FCI, so by the time we reach the CCSDT or CCSDTQ levels, the recovery of the correlation energy is virtually perfect (see Figure 1.1). This should be contrasted with truncated CI and finite-order MBPT approaches which converge much slower or (in the MBPT case) erratically. Figure 1.1 gives the impression that CISDTQ catches up with CCSDTQ, having similar computational costs, but this is only true for small molecules used to generated Figure 1.1, for which size inextensivity of CI is a small problem. In the thermodynamic limit, truncated CI approaches recover no correlation energy.
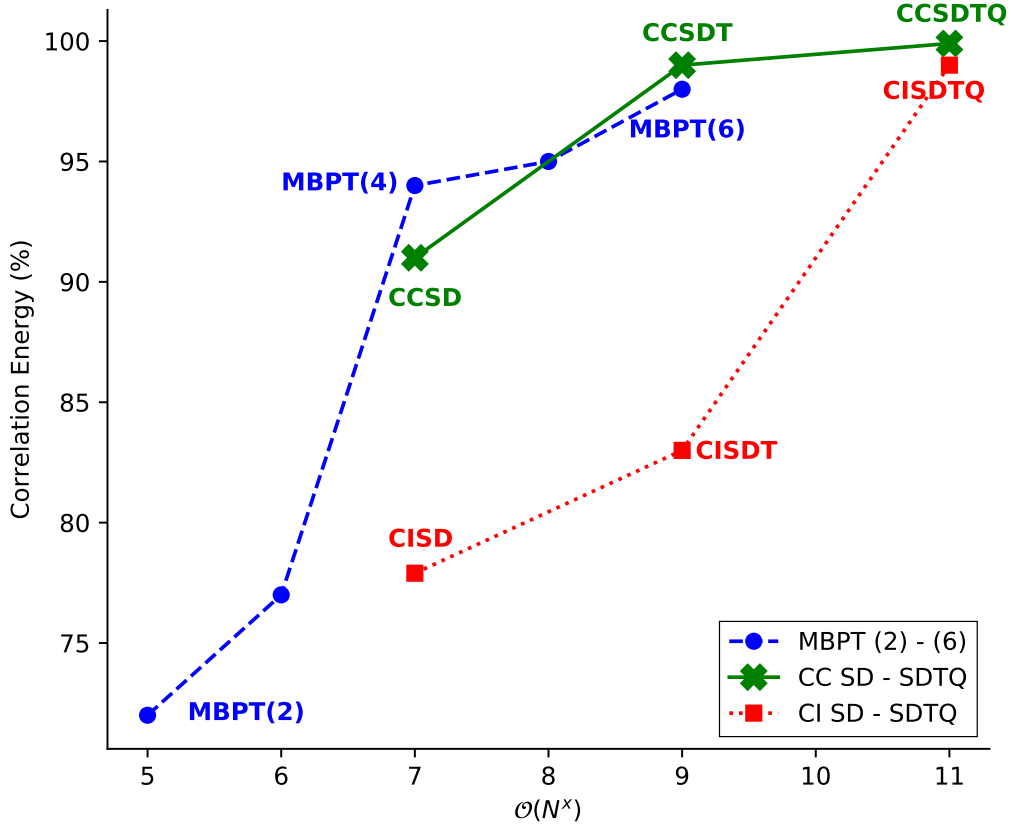
Figure 1.1: Plot comparing the percentage of correlation energy captured by three different classes of QC methods. The x-axis represents a measure of the CPU time scaling for the various approximations within each family. Adapted from Ref. [42].

Having several desirable features, the conventional approximations based on the CC theory employing a single reference determinant are not exempt of problems. This is particularly true when nondynamical electron correlation effects become large and strongly correlated systems are examined [43–49]. For example, the basic CCSD approach, which is nowadays routinely applicable to systems consisting of dozens of non-hydrogen atoms and hundreds of correlated electrons, fails in situations of stronger multi-reference correlations, such as those found in bond breaking phenomena and systems having substantial biradical or polyradical [50, 51]. For strongly correlated systems, such as those encountered when examining metal-insulator transitions or low-dimensional metalic species, the situation may become so dramatic that CCSD becomes

singular and real solutions may not exist [43–49]. To address these issues, two fundamentally different approaches can be exploited within the CC framework. The first approach defining the multi-reference (MR) CC methods, introduces a multi-dimensional model space, represented with a set of reference Slater determinants, that provides an appropriate zeroth-order description of the quasi-degenerate electronic states of interest. This set is typically obtained by means of complete-active-space (CAS) SCF calculations, which capture most of the relevant nondynamical correlation effects, and which are extensively used in other kinds of MR methods (e.g., MRCI and MRMBPT). Once the multi-dimensional model space is defined, which is typically done with the help of active orbitals, the remaining correlations are captured using one of the suitable generalizations of the CC formalism to MR situations (see Refs. [42, 52] for reviews). An example is the Jeziorski–Monkhorst wave function ansatz [53] that introduces multiple cluster operators associated with each of the reference determinants spanning the model space. This ansatz has been exploited within a multi-root, state-universal, description [53–76] and, in recent years, in a state-specific manner [77–84]. The second possibility is to take advantage of the aforementioned single-reference CC (SRCC) wave function ansatz, $|\Psi_0\rangle = e^T |\Phi\rangle$, and its equation-of-motion (EOM) extension to excited [85–92], electron attached, and ionized states [93–106], $|\Psi_\mu\rangle = R_\mu e^T |\Phi\rangle$, where R_μ is a suitable electronic excitation, electron attaching, or ionizing operator leading to the excited-state wave functions $|\Psi_\mu\rangle$ ($\mu > 0$). In these cases, instead of using multiple reference determinants to capture nondynamical correlations *a priori*, one recovers them, along with the corresponding dynamical correlation effects, dynamically, i.e., through the presence of higher-than-two-body components of the T and R_μ operators acting on a single reference determinant. In this manner, we retain the much appreciated conceptual simplicity of the SRCC methodology, unparalleled by the much more complex MRCC approaches. Nevertheless, by resorting to a single reference, we must often rely on the higher-level SRCC or EOMCC methods that include singles, doubles, and triples

(CCSDT/EOMCCSDT) [36, 37, 88, 89, 107–109], or even quadruples (CCSDTQ/EOM-CCSDTQ) [38, 40, 41, 107, 108, 110, 111], pentuples (CCSDTQP/EOMCCSDTQP) [112] and hextuples (CCSDTQPH/EOMCCSDTQPH) [107, 113–117], in order to overcome the failures in describing higher-order correlation effects appearing in multi-reference situations. This creates an enormous challenge, since the full CCSDT/EOMCCSDT and CCSDTQ/EOMCCSDTQ methods have iterative CPU time steps that scale as $n_o^3 n_u^5$ and $n_o^4 n_u^6$, respectively (n_o and n_u are the numbers of correlated occupied and unoccupied orbitals in a basis), with methods beyond quadruples being even more expensive, limiting approaches of this type to small molecular problems with a dozen or so correlated electrons. As a consequence, and following the trend in QC toward the best balance between accuracy and computational effort, much work in the SRCC and EOMCC theories has been focused on incorporating higher-than-two-body components of T and R_μ operators in an approximate and tractable manner.

Historically, the commonly adopted procedure to estimate the contributions of the higher-order components of T and R_μ has been to employ MBPT arguments. This has led to the development of the iterative CCSDT-n [45, 118–123] and CCSDTQ-n [124] schemes and their non-iterative CCSD[T] [125], CCSD(T) [126], and CCSD(TQ$_\mathrm{f}$) [127] counterparts as well as their various excited-state extensions based on EOMCC and linear response CC [128] (see, e.g., [129–134] on CC3, EOMCCSD(T), etc.). Methods in this category reduce the prohibitive costs of the full CCSDT/EOMCCSDT and CCSDTQ/EOMCCSDTQ computations, recovering some of the mostly dynamical correlations missing in CCSD, but are still incapable of describing stronger nondynamical correlations and quasi-degenerate electronic states, due to their divergent MBPT nature. Therefore, in order to tackle systems characterized by this kind of stronger correlations, one has to shift away from traditional MBPT concepts and turn to other ways of estimating higher-than-two-body components of T and R_μ.

Among the most successful approaches in this area are the completely renormal-

ized (CR) CC/EOMCC schemes, resulting from the formalism of the method of moments of CC equations (MMCC) [51, 135–152], and the active-space CC/EOMCC theories [41, 50, 51, 88–92, 101–106, 153–164]. The former approaches are based on the idea of correcting the energies obtained in the lower-order CC/EOMCC calculations, such as CCSD or EOMCCSD, using *a posteriori* state-specific and non-iterative corrections δ_μ ($\mu \geq 0$), designed with the help of moment energy expressions. A few examples of methods in this category are CR-CCSD(T) [135–137, 149, 150], CR-CCSD(TQ) [135–137, 149, 150], CR-CC(2,3) [144–147], CR-CC(2,4) [144, 145], CR-CC(2,3)+Q [165, 166], and their corresponding EOM analogs, such as CR-EOMCCSD(T) [140], δ-CR-EOMCCSD(T) [140], CR-EOMCC(2,3) [146, 148], and δ-CR-EOMCC(2,3) [151, 167]. Methods of this type work well in situations involving electronic quasi-degeneracies, such as single (triples correction) and double (quadruples correction) bond breaking, selected biradicals, and excited states dominated by one- as well as two-electron transitions, especially when compared to the previously discussed MBPT-based CC/EOMCC approximations. Nevertheless, they cannot be applied to all MR situations, such as biradicals involving delocalized π networks and reaction mechanisms where biradicals of this type appear [163, 164], largely because of the intrinsic *a posteriori* nature of the non-iterative corrections δ_μ where, in analogy to conventional MBPT-based approaches, such as CCSD(T), there is no mechanism to relax lower-order components of T and R_μ in the presence of their higher-order counterparts. For example in methods such as CR-CC(2,3), CR-CC(2,4), or CR-EOMCC(2,3), which are used to correct the CCSD or EOMCCSD energies for triples or triples and quadruples, it is impossible for the one- and two-body components of T and R_μ to relax in the presence of their higher-order counterparts, which may lead to inaccurate values of singly and doubly excited amplitudes that propagate into corrections δ_μ, diminishing their usefulness.

This issue can be successfully counteracted if we switch to the active-space CC/EOMCC approaches [41, 88–90, 153–162, 168, 169], which nowadays exist (following

9

[88–90, 160, 161]) under the acronyms such as CCSDt/EOMCCSDt, CCSDtq/EOMCCS-Dtq, CCSDTq/EOMCCSDTq, etc., where one uses a small subset of active orbitals to select higher-than-two-body components of T and R_μ operators, such as triples ("little" t), quadruples ("little" q), or others, within the parent CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, etc. frameworks. As a result, the active-space CC/EOMCC methods, such as CCSDt/EOMCCSDt and CCSDtq/EOMCCSDtq, provide great results in MR situations, including, but not restricted to, single and double bond breaking, reaction pathways involving biradicals, and excited states dominated by two-electron transitions, at a small fraction of the computational costs of the parent full CCSDT/EOMCCSDT and CCSDTQ/EOMCCSDTQ approaches. However, by adopting the active-space methodology, one loses the efficiency in capturing dynamical correlations, which characterizes the CCSD(T), CR-CC(2,3), and similar approaches. This issue can become severe when studying more complicated systems, such as the aforementioned reaction pathways involving biradical transition states characterized by delocalization of π electron networks, as in the automerization of cyclobutadiene and the isomerization of bicyclobutane to butadiene, where, CCSDt misses about 10–20 mE_h of mostly dynamical correlation energy from the reactant, product, and transition states [164] compared to its CCSDT parent, which may introduce 2–3 kcal/mol errors in relative energies. Moreover, the advantage of the purely black-box nature of the standard CC/EOMCC formalism is partially lost because of the user- and system-dependent selection of active orbitals required in these approaches.

The limitations in the existing CR-CC/CR-EOMCC and active-space CC/EOMCC approaches, namely, the absence of the coupling among the lower- and higher-order components of T and R_μ operators in the former methods and the loss of information about certain categories of higher-order dynamical correlation effects in the latter, can be naturally alleviated if we merge both classes of methods. Doing so gave birth to the CC(P;Q) framework [51, 163, 164, 170], which has the potential to address the above challenges in

an elegant, robust, and computationally efficient manner. The key idea of the CC(P;Q) formalism is to perform CC/EOMCC calculations in a subspace of the many-electron Hilbert space spanned by Slater determinants that dominate the ground or ground and excited states of interest, called the P space, which are subsequently corrected for the correlation effects described by the determinants in a complementary subspace, called the Q space. This is achieved by using non-iterative corrections δ_μ based on moment energy expansions similar to those exploited in MMCC. If the Q space is the orthogonal complement of the P space and all elements of the moment expansions are treated fully, the CC(P;Q) considerations result in the exact, FCI description. Otherwise we end up with approximations, which we can classify as conventional and unconventional. An example of a conventional choice would be the P space spanned by singly and doubly excited determinants and the Q space spanned by triply excited determinants. In that case the CC(P;Q) calculations become equivalent to the aforementioned CR-CC(2,3) or CR-EOMCC(2,3) schemes, where one uses δ_μ corrections to correct the CCSD or EOMCCSD energies. Similarly, the CR-CC(2,4) or CR-EOMCC(2,4) methods are obtained by letting the P space be spanned by singly and doubly excited determinants and the Q space by triply and quadruply excited determinants. However, the great potential of the CC(P;Q) framework lies in the fact that one can also make other, unconventional choices of the P and Q spaces to further improve the results when conventional choices are too constraining. We can, for example, choose the P space to be spanned by all singles and doubles and a subset of triples or triples and quadruples defined via active orbitals, as in CCSDt/EOMCCSDt or CCSDtq/EOMCCSDtq, to capture the coupling among the lower and leading higher-order components of the T and R_μ operators, and define the Q space as a span of the remaining triples or the remaining triples and quadruples, to capture those, predominantly dynamical, correlations that may be missing in CCSDt/EOMCCSDt and CCSDtq/EOMCCSDtq. This leads to the hierarchy of CC(P;Q) methods, introduced in Refs. [51, 163, 164, 170] and abbreviated as CC(t;3), CC(t,q;3), CC(t,q;3,4), etc., where the

letter symbols explain what higher-than-double excitations are included in the P-space CC/EOMCC calculations (with 't' representing the active-space triples of CCSDt/EOM-CCSDt and 't,q' the active-space triples and quadruples of CCSDtq/EOMCCSDtq) and the integers designate the many-body ranks of the missing excitations included in the Q space corrections δ_μ (with '3' representing missing triples; '4', missing quadruples, etc.). Several benchmark calculations, including single and double bound dissociations, reaction pathways involving biradical transition states, singlet–triplet gaps in biradical and prototype magnetic species, and noncovalent interactions in challenging alkaline earth metal dimers, reported in Refs. [163, 164, 164, 170–173], demonstrated that the CC(t;3), CC(t,q;3), and CC(t,q;3,4) approaches are capable of reproducing the total and relative energies obtained using their respective CCSDT and CCSDTQ parent methods, to within small fractions of a millihartree at much reduced computational costs. Forth-coming publications are in preparation in our group that will show that the same is true when applying the CC(P;Q)-based CC(t;3) approach to excited-state potential energy surfaces along bond breaking coordinates.

All of the above demonstrates that the CC(P;Q) methodology of Refs. [51, 164, 170] is a powerful formalism, which provides the opportunity to develop efficient and robust SR-CC-type methods that can accurately approximate energetics of higher-level CC approaches, such as CCSDT or CCSDTQ, at much more affordable computational costs, even when electronic quasi-degeneracies become substantial. However, as presently implemented, the existing CC(P;Q)-based CC(t;3), CC(t,q;3) or CC(t,q;3,4) methods are not purely black-box, since one has to manually select user- and system-dependent active occupied and active unoccupied orbitals to define the underlying P-space CC(P) computations. The following questions emerge: (i) Is there an automated way of determining P spaces reflecting on the nature of states being calculated, while using Q-space corrections δ_μ to capture the remaining correlations of interest?; (ii) Can this be done such that the resulting electronic energies rapidly converge to their high-level (CCSDT, CCSDTQ,

etc.) parents, even in quasi-degenerate situations where higher-than-two-body components of T (and R_μ) become large, at the small fraction of the computational cost and with an ease of a black-box calculation? One of the successes of this thesis project is the demonstration that both questions have positive answers if we combine the deterministic CC(P;Q) methodology with the stochastic CI (and CC) Quantum Monte Carlo (QMC) ideas. To continue, let us first introduce the CIQMC (and the analogous CCMC) approaches.

The QMC approaches are based on the original ideas that date back to the 1949 work by Metropolis and Ulam [174] of numerically integrating the Fokker–Planck, Schrödinger, and similar differential equations with the help of stochastic sampling [174–176]. This procedure gave birth to the now ubiquitous variational Quantum Monte Carlo (VMC) method [177, 178], in which parameters of a trial wave function are optimized variationally, by relying on the stochastic computation of the required expectation values of the Hamiltonian operator. Later, the diffusion Quantum Monte Carlo (DMC) approaches [179–181] were introduced, which improve upon the VMC methodology by allowing a trial wave function to "diffuse" onto the exact wave function by adopting a projection technique (cf. Refs. [182–184] for further remarks and details). Over the years, DMC approaches have proven to be useful and often very accurate. This is partly because they operate directly on the real space of $3N$ particle coordinates within a probability distribution defined by physical laws, which allows them to bypass the issues arising from using finite basis sets, while avoiding biases related to *a priori* assumptions regarding dominant wave function components that one usually has to make in conventional QC. The DMC (also VMC) method are also easy to parallelize across many multi-core nodes, making them perfect for the efficient use of modern computer architectures, which is a challenging technical obstacle afflicting a large fraction of currently available QC methods using basis sets. Despite these impressive advantages, DMC methods suffer from a major issue. The DMC propagations by themselves are unaware of the nodal structure

of the target anti-symmetric many-fermion wave function, which is a key requirement in order to satisfy Pauli's exclusion principle. If this problem is left untreated, the DMC calculations will converge to nodeless totally symmetric bosonic states, defeating the purpose of the computation entirely. A common way of ameliorating this so-called "boson catastrophe" or "fermion sign problem" is to employ the fixed-node approximation [185–188], which consists in using less-expensive deterministic QC methods to obtain an approximate nodal structure of the target wave function, *a priori*, and enforce it during the QMC simulations. Unfortunately, finding correct nodal structures in the highly dimensional manifolds characterizing electronic and other many-fermion wave functions is a non-trivial task. The less expensive methods of QC may not be up to it, resulting in loss of accuracy in DMC computations.

In response to this issue, a QMC method has emerged that circumvents the fixed-node approximations altogether at the expense of returning to the second-quantized many-electron Hilbert space spanned by Slater determinants constructed from single-particle basis functions. This new approach is called FCIQMC and is designed to converge the exact, FCI, wave function in the limit of infinite imaginary propagation time by employing a walker population dynamics algorithm, reminiscent of DMC, that explores the space of all possible N-electron Slater determinants by sampling their corresponding weights [189]. This method has since been improved via the initiator algorithm to accelerate convergence [1, 190], and also extended to excited states [191], solids [192], spin-adapted Hamiltonians [193], and, more recently, to exploit Jastrow-factorized explicitly correlated wave functions [194–197]. The idea of a QMC algorithm exploiting the space of Slater determinants has also been employed with an alternative parametrization of the FCI wave function, namely, the CC ansatz. The resulting CCMC methods introduced by Thom in [198], have also adopted the initiator algorithm [199] and further improved by directly sampling the similarity transformed Hamiltonian [200]. More recently, the CCMC framework has been extended to include MRCC wave functions

[201] using the ideas of splitting the cluster operator into internal and external parts introducing in Ref. [158], which is a basis of active-space CC methods.

Returning to the question presented earlier on the non-black-box behavior of the active-space-based CC(P;Q) methods, one of the main objectives of this dissertation is to show that the FCIQMC methodology and its truncated CIQMC and CCMC counterparts can be successfully merged with the general CC(P;Q) framework to provide a black-box, effective, and automatic solution to the selection of the appropriate P and Q subspaces of the many-electron Hilbert space [202, 203]. This work addresses this merger by capitalizing on wave functions acquired using CIQMC (and CCMC) propagations. By identifying their leading determinantal components, we can immediately and efficiently construct suitable P spaces, which allow for the relaxation of lower-order cluster components in the presence of higher-order ones. Moreover, by selecting the excited determinants not captured by the corresponding initiator CIQMC (*i*-CIQMC) propagations, we can define Q spaces capable of correcting for the missing mostly dynamical correlation effects using CC(P;Q) moment corrections δ_μ. As I will demonstrate in the body of this document, the new semi-stochastic CC(P;Q) formalism is capable of accurate reproducing the energetics corresponding to the parent CCSDT, CCSDTQ, and EOMCCSDT methods out of the early stages of CIQMC propagations without manual intervention needed in the active-space approaches, while reducing the computational cost associated with such high-level calculations by one or more orders of magnitude.

The second goal of this dissertation is the introduction of yet another method combining the strengths of the deterministic CC theory with the stochastic FCIQMC, this time aiming at recovering the exact, FCI energetics [204]. This new method, which we call cluster-analysis-driven (CAD) FCIQMC, is based on the observation that exact, FCI ground-state energies can be obtained by solving CCSD-like equations in the presence of exact triply and quadruply excited cluster components. Historically, this idea has motivated the work on the so-called externally corrected CC methods [43–49, 118, 205–214],

which take advantage of approximate non-CC sources of triply and quadruply excited cluster components that have a more robust physical behavior under the particular circumstances of the chemical system of interest (e.g., bond breaking, strong correlations, etc.). I will demonstrate in this thesis that the initiator-approximation-based i-FCIQMC method [1] is a great candidate for providing such high quality triply and quadruply cluster components allowing us to substantially accelerate convergence of the purely stochastic i-FCIQMC propagations toward FCI with the help of the semi-stochastic CAD-FCIQMC computations, especially in a bond stretching or breaking regime.

CHAPTER 2

THEORY

The first step in the description of any QC method is to write the N-electron Schrödinger equation. In this thesis, we are interested in its Born–Oppenheimer approximation, which simplifies matters by assuming that the electron coordinates, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, are decoupled from the position of the nuclei, $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_M)$, such that the resulting many-electron wave function depends parametrically on $\mathbf{R}$. Such assumption can be understood as a model, in which electrons interact with the nuclei as if the latter were nothing more than static point charges, and arises from the fact that electrons move many orders of magnitude faster than the nuclei. Within these considerations, the resulting electronic Schrödinger equation becomes

$$H\Psi_\mu(\mathbf{X}; \mathbf{R}) = E_\mu(\mathbf{R})\Psi_\mu(\mathbf{X}; \mathbf{R}), \tag{2.1}$$

where $E_\mu(\mathbf{R})$ and $\Psi_\mu(\mathbf{X}; \mathbf{R})$ are the energy and many-electron wave function corresponding to the state $\mu \geq 0$. H is the Hamiltonian operator, consisting of one- and two-body terms, such that

$$H = \sum_{i=1}^{N} z(\mathbf{x}_i) + \sum_{i<j}^{N} v(\mathbf{x}_i, \mathbf{x}_j), \tag{2.2}$$

and

$$z(\mathbf{x}_i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{R_{iA}}, \quad v(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{r_{ij}}, \tag{2.3}$$

where $z(\mathbf{x}_i)$ describes the kinetic energy of the i-th electron as well as its electron-nuclei attraction potential, and $v(\mathbf{x}_i, \mathbf{x}_j)$, the inter-electronic repulsion between electrons i and j, in atomic units. In these expressions, N is the number of electrons, $\mathbf{x}_i = (\mathbf{r}_i, \sigma_i)$ are the coordinates of the i-th electron, such that $\mathbf{r}_i = (x_i, y_i, z_i)$ and σ_i is the spin coordinate, M is the number of nuclei in the system, Z_A is the nuclear charge of the A-th nucleus, R_{iA} is the distance between the i-th electron and A-th nucleus, and r_{ij} is the distance between electrons i and j.
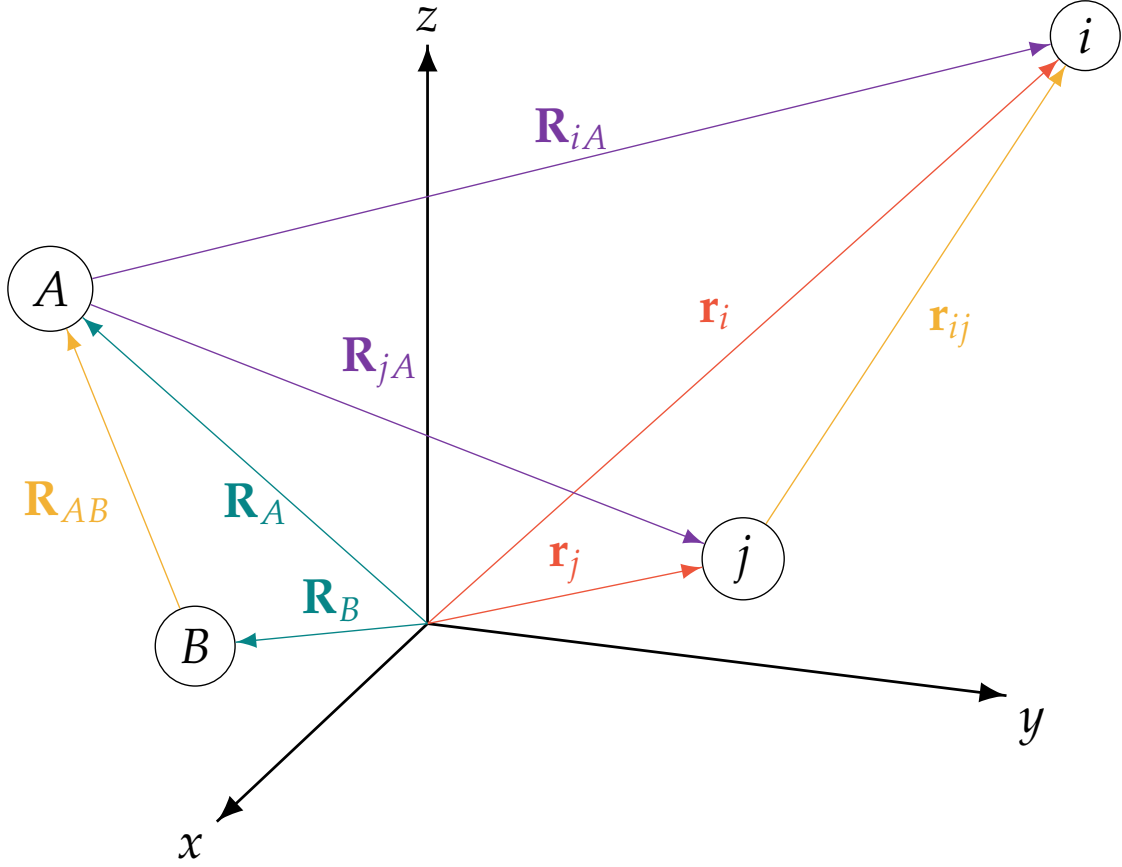
Figure 2.1: A Cartesian representation of the molecular coordinate system, where indices i, j represent electrons and A, B, nuclei. The vectors $\mathbf{r}_i$ and $\mathbf{r}_j$ denote the Cartesian components of the electron coordinates, and $\mathbf{R}_A$ and $\mathbf{R}_B$ are the analogous vectors defining the positions of nuclei.

In order to obtain the second-quantized expressions for the one- and two-body operators, one must first introduce an abstract and complete system of discrete and orthonormal one-particle states, $|p\rangle$, defined for a single electron. These basis states, that together form a single-particle Hilbert space, $\mathscr{V}_1$, are represented in the real coordinate space by

$$\varphi_p(\mathbf{x}_i) = \langle \mathbf{x}_i | p \rangle. \tag{2.4}$$

With this expression, and the resolution of identity given by the relation $\mathbb{1} = \int d\mathbf{x}_i \, |\mathbf{x}_i\rangle \langle \mathbf{x}_i|$, one can conclude that for any pair of states, $|p\rangle$ and $\langle q|$ (where $\langle q| = |q\rangle^\dagger$),

the inner product between them is given by the following cases:

$$\langle p|q \rangle = \delta_{pq}, \quad \delta_{pq} = \begin{cases} 1, & \text{for } p = q, \\ 0, & \text{for } p \neq q \end{cases}, \tag{2.5}$$

where δ_{pq} is the Kronecker delta, since

$$\langle p|q \rangle = \int d\mathbf{x}_i \, \langle p|\mathbf{x}_i \rangle \, \langle \mathbf{x}_i|q \rangle = \int d\mathbf{x}_i \, \varphi_p^*(\mathbf{x}_i) \varphi_q(\mathbf{x}_i) = \delta_{pq}. \tag{2.6}$$

In this work, we are interested in many-electron systems. Therefore, we must also introduce a basis of N-electron states spanning the many-body Hilbert space and satisfying the Pauli exclusion principle. This is achieved by taking the tensor product of single-particle spaces, so that the N-electron Hilbert space becomes $\mathcal{H}^N = \otimes_{i=1}^{N} \mathcal{V}_1$, and retaining the corresponding anti-symmetric sub-space, such that an appropriately normalized N-electron basis function is given by

$$|\{p_1 \ldots p_N\}\rangle = \sqrt{N!} \mathcal{A} \, |p_1\rangle \otimes \cdots \otimes |p_N\rangle, \tag{2.7}$$

where $|p_1\rangle, \ldots, |p_N\rangle \in \mathcal{V}_1$. In this expression, the antisymmetrizer,

$$\mathcal{A} = (N!)^{-1} \sum_{P \in \mathcal{S}_N} (-1)^P P, \tag{2.8}$$

consists of all permutations, P, elements of the symmetric group of order N, $\mathcal{S}_N$, along with the signs, $(-1)^P$, resulting from their corresponding permutation parities, ρ. The functions, $|\{p_1 \ldots p_N\}\rangle$, are called Slater determinants and are the basis states of the N-body Hilbert space $\mathcal{H}^N$. To avoid repetitions of equivalent determinants, which result from permutations among the same single-particle states, such that

$$|\{p_{\sigma_1} \ldots p_{\sigma_N}\}\rangle = (-1)^\rho \, |\{p_1 \ldots p_N\}\rangle, \tag{2.9}$$

are two determinants related only by a phase factor, $(-1)^\rho$, where the indices $p_{\sigma_1} \ldots p_{\sigma_N}$ represent a particular permutations of the indices $p_1 \ldots p_N$, the Slater determinants $|\Phi_k\rangle = |\{p_1 \ldots p_N\}\rangle$ are subjected to a particular single-particle ordering. Usually, and

19

throughout this work, they are organized such that the single-particle indices obey the ascending ordering given by $p_1 < p_2 < \cdots < p_n$, but other organization schemes are also possible as long as only linearly independent determinants are kept in the many-electron basis set. The resulting uniquely ordered determinants, $|\Phi_k\rangle$, form a complete set of basis functions that satisfy orthonormality, $\langle\Phi_k|\Phi_l\rangle = \delta_{kl}$, as can be proven from Eqs. (2.6) and (2.7). In analogy to the single-particle basis, the N-particle states, $|\Phi_k\rangle$, can be represented in the coordinate space as follows:

$$\Phi_k(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \langle\mathbf{x}_1, \ldots, \mathbf{x}_N|\Phi_k\rangle = \sqrt{N!}\mathscr{A}\,\varphi_{p_1}(\mathbf{x}_1)\ldots\varphi_{p_N}(\mathbf{x}_N). \tag{2.10}$$

Furthermore, the creation and annihiliation operators, a^p and a_q, associated with the indices p and q, are defined as abstract operators in the more general Fock space, $\mathcal{F} = \bigoplus_{n=0}^{\infty} \mathscr{H}^n$, consisting of the direct sum of all n-electron Hilbert spaces, where $n \in \{0, 1, 2, \ldots\}$. In this larger space, which accepts non-particle-conserving operations, the operators a^p and a_q respectively increase and decrease the number of particles in a given Hilbert space, $\mathscr{H}^N$, by one. For example, in the equalities,

$$a_q\,|\{qp_2\ldots p_N\}\rangle = |\{p_2\ldots p_N\}\rangle \tag{2.11}$$

and

$$a_q\,|\{p_2\ldots p_N\}\rangle = 0, \quad \forall\,q \neq p_2, \ldots, p_N, \tag{2.12}$$

a_q annihilates the electron occupying the single-particle state q in $|\{qp_2\ldots p_N\}\rangle \in \mathscr{H}^N$ producing $|\{p_2\ldots p_N\}\rangle \in \mathscr{H}^{N-1}$, with both states belonging to the Fock space. Similarly, as shown in

$$a^p\,|\{p_2\ldots p_N\}\rangle = |\{pp_2\ldots p_N\}\rangle \tag{2.13}$$

and

$$a^p\,|\{pp_2\ldots p_N\}\rangle = 0, \quad \forall\,p \neq p_2, \ldots, p_N, \tag{2.14}$$

a^p creates an electron in the single-particle state p which is unoccupied in $|\{p_2\ldots p_N\}\rangle \in \mathscr{H}^{N-1}$ yielding $|\{pp_2\ldots p_N\}\rangle \in \mathscr{H}^N$. Since these states represent many-electron functions, which obey fermionic statistics and, thus, satisfy Pauli's exclusion principle, a

creation operator cannot create a particle already present in a Slater determinant and its annihilation counterpart must not remove an already absent one. Such illegal operations will always result in zero, regardless of the permutations or phase factors associated with a Slater determinant.

Now that we have introduced the fundamental elements of the second quantized theory, let us proceed to define general one- and two-body operators, O_1 and O_2, using this formalism. To do so, we begin by writing their general definitions in the real coordinate space:

$$O_1(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \sum_{i=1}^{N} o_1(\mathbf{x}_i), \tag{2.15}$$

and

$$O_2(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \sum_{i<j}^{N} o_2(\mathbf{x}_i, \mathbf{x}_j). \tag{2.16}$$

Using the identities given by Eq. (2.4) and

$$f(\mathbf{x}_i) = \sum_p \int d\mathbf{x}_i' \varphi_p^*(\mathbf{x}_i') f(\mathbf{x}_i') \varphi_p(\mathbf{x}_i), \tag{2.17}$$

one can easily arrive at expressions defining the matrix elements of $o_1(\mathbf{x}_i)$,

$$\langle p|o_1|q\rangle = \int d\mathbf{x}_1 \, \varphi_p^*(\mathbf{x}_1) \, o_1(\mathbf{x}_1) \, \varphi_q(\mathbf{x}_1), \tag{2.18}$$

and $o_2(\mathbf{x}_i, \mathbf{x}_j)$

$$\langle pq|o_2|rs\rangle = \iint d\mathbf{x}_1 \, d\mathbf{x}_2 \, \varphi_p^*(\mathbf{x}_1)\varphi_q^*(\mathbf{x}_2) \, o_2(\mathbf{x}_1, \mathbf{x}_2) \, \varphi_r(\mathbf{x}_1)\varphi_s(\mathbf{x}_2), \tag{2.19}$$

where the indices p, q, r, and s, are dummy labels for states in $\mathcal{V}_1$.

By expanding the determinant $\Phi_k(\mathbf{x}_1,\ldots,\mathbf{x}_N)$ in terms of its single-particle function components, using Eq. (2.10), and employing the identities described above, we obtain a second quantization expression for a general one-body operator in the N-electron Hilbert space, $O_1(\mathbf{x}_1,\ldots,\mathbf{x}_N)$, acting on a determinant $\Phi_k(\mathbf{x}_1,\ldots,\mathbf{x}_N)$, which is given by

$$O_1(\mathbf{x}_1,\ldots,\mathbf{x}_N) \, \Phi_k(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \sum_{pq} \langle p|o_1|q\rangle \, a^p a_q \, |\Phi_k\rangle. \tag{2.20}$$

Analogously, a general two-body operator, $O_2(\mathbf{x}_1, \ldots, \mathbf{x}_N)$, acting on N-electron functions in the Hilbert space can be expressed in the language of second quantization using the following expression:

$$O_2(\mathbf{x}_1, \ldots, \mathbf{x}_N)\, \Phi_k(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \frac{1}{2} \sum_{pqrs} \langle pq|o_2|rs \rangle\, a^p a^q a_s a_r\, |\Phi_k\rangle. \tag{2.21}$$

Using these expressions for the general one- and two-body operators, and employing the Einstein summation convention, in which repeated upper and lower indices are summed over, we can rewrite the one- and two-body components of the electronic Hamiltonian, which become

$$Z = \sum_{pq} \langle p|z|q \rangle\, a^p a_q = z_p^q\, a^p a_q, \tag{2.22}$$

and

$$V = \frac{1}{2} \sum_{pqrs} \langle pq|v|rs \rangle\, a^p a^q a_s a_r = \frac{1}{2} v_{pq}^{rs}\, a^p a^q a_s a_r. \tag{2.23}$$

During the remaining chapters and sections of this work, we will use these expressions to define the Hamiltonian operator.

2.1 Coupled-Cluster Theory and its Equation-of-Motion Extension to Excited States

As already mentioned in the Introduction, all methods of the SRCC theory exploit the exponential wave function ansatz, given by

$$|\Psi_0^{(A)}\rangle = e^{T^{(A)}}\, |\Phi\rangle. \tag{2.24}$$

In the above expression, the cluster operator $T^{(A)}$ is the operator T truncated at a conveniently chosen m_A-tuply excited component, T_{m_A}, which gives rise to the conventional CC approximation denoted as method A so that when $A = \text{CCSD}$, $m_A = 2$, when $A = \text{CCSDT}$, $m_A = 3$, and when $A = \text{CCSDTQ}$, $m_A = 4$. In general,

$$T^{(A)} = \sum_{n=1}^{m_A} T_n, \tag{2.25}$$

and

$$T_n = \sum_{\substack{i_1 < \cdots < i_n \\ a_1 < \cdots < a_n}} t_{a_1 \ldots a_n}^{i_1 \ldots i_n} E_{i_1 \ldots i_n}^{a_1 \ldots a_n}, \tag{2.26}$$

is the n-body component of $T^{(A)}$ defined in terms of the antisymmetrized cluster amplitudes $t_{a_1 \ldots a_n}^{i_1 \ldots i_n}$ and the elementary excitation operators $E_{i_1 \ldots i_n}^{a_1 \ldots a_n}$ promoting electrons occupying spin-orbitals $i_1, \ldots, i_n$ in a reference determinant $|\Phi\rangle$ to the unoccupied spin-orbitals $a_1, \ldots, a_n$, as introduced by Čížek in [25].

Formally, we define

$$E_{i_1 \ldots i_n}^{a_1 \ldots a_n} = \prod_{k=1}^{n} a^{a_k} a_{i_k}, \tag{2.27}$$

where a^{a_k} (a_{i_k}) is the creation (annihilation) operator, associated with the unoccupied, a_k, (occupied, i_k) spin-orbital.

The SRCC theory can further be extended to include excited electronic states, $|\Psi_\mu\rangle$ ($\mu > 0$), by employing the EOMCC formulation

$$|\Psi_\mu^{(A)}\rangle = R_\mu^{(A)} |\Psi_0^{(A)}\rangle = R_\mu^{(A)} e^{T^{(A)}} |\Phi\rangle, \tag{2.28}$$

in which we apply the linear excitation operator

$$R_\mu^{(A)} = r_{\mu,0} \mathbb{1} + \sum_{n=1}^{m_A} R_{\mu,n}, \tag{2.29}$$

to the ground CC state defined in Eq. (2.24). Again, we use the symbol A to denote the various truncations in T and R_μ in order to be consistent with Eq. (2.24). Also,

$$R_{\mu,n} = \sum_{\substack{i_1 < \ldots < i_n \\ a_1 < \ldots < a_n}} r_{\mu,a_1 \ldots a_n}^{i_1 \ldots i_n} E_{i_1 \ldots i_n}^{a_1 \ldots a_n}, \tag{2.30}$$

defines the n-body components of $R_\mu^{(A)}$.

In analogy to the ground-state case, different choices of m_A result in various conventional EOMCC approximations. For example, $m_A = 2$, $m_A = 3$, and $m_A = 4$ correspond to the EOMCCSD, EOMCCSDT, and EOMCCSDTQ approaches.

With the above definitions of the ground-state CC and excited-state EOMCC wave functions, we proceed with the determination of the cluster and linear excitation amplitudes, $t_{a_1 \ldots a_n}^{i_1 \ldots i_n}$ and $r_{\mu, a_1 \ldots a_n}^{i_1 \ldots i_n}$, respectively. In the ground-state case, we begin by rewriting the corresponding eigenvalue problem, $H |\Psi_0\rangle = E_0 |\Psi_0\rangle$, defining the Schrödinger equation, such that $|\Psi_0\rangle$ is approximated by $|\Psi_0^{(A)}\rangle$. If we multiply the resulting expression by $e^{-T^{(A)}}$, and introduce the similarity-transformed Hamiltonian, $\bar{H}^{(A)} = e^{-T^{(A)}} H e^{T^{(A)}} = (H e^{T^{(A)}})_C$, where the subscript C indicates the connected part of the corresponding operator product, we obtain the connected-cluster form of the Schrödinger equation

$$\bar{H}^{(A)} |\Phi\rangle = E_0^{(A)} |\Phi\rangle . \tag{2.31}$$

Next, we project this expression, onto all excited determinants $|\Phi_{i_1 \ldots i_n}^{a_1 \ldots a_n}\rangle = E_{i_1 \ldots i_n}^{a_1 \ldots a_n} |\Phi\rangle$ with $n \leq m_A$, obtained from all possible combinations of occupied, $i_1, \ldots, i_n$, and unoccupied, $a_1, \ldots, a_n$ orbitals. This results in a system of non-linear, energy-independent equations

$$\langle \Phi_{i_1 \ldots i_n}^{a_1 \ldots a_n} | \bar{H}^{(A)} | \Phi \rangle = 0, \quad i_1 < \cdots < i_n, \quad a_1 < \cdots < a_n, \quad n = 1, \ldots, m_A, \tag{2.32}$$

which are solved in an iterative fashion, with the goal of determining the converged cluster amplitudes $t_{a_1 \ldots a_n}^{i_1 \ldots i_n}$ to within an *a priori* prescribed threshold. Following this step, we calculate the ground-state SRCC energy, $E_0^{(A)}$, directly, by taking the expectation value of the similarity transformed Hamiltonian,

$$E_0^{(A)} = \langle \Phi | \bar{H}^{(A)} | \Phi \rangle , \tag{2.33}$$

where the t amplitudes entering $\bar{H}^{(A)}$ have been replaced by the converged ones.

Once the the ground-state amplitudes $t_{a_1 \ldots a_n}^{i_1 \ldots i_n}$ and the corresponding energy $E_0^{(A)}$ are computed, we can proceed toward the determination of the excited-state electronic energies, $E_\mu^{(A)}$, and linear-excitation amplitudes, $r_{\mu, a_1 \ldots a_n}^{i_1 \ldots i_n}$, where $\mu > 0$. To do so, we must first insert the EOMCC wave function ansatz given by Eq. (2.28) into the electronic

24

Schrödinger equation, $H |\Psi_\mu\rangle = E_\mu |\Psi_\mu\rangle$, and multiply the resulting equation by $e^{-T^{(A)}}$ from the left. After projecting this expression onto all excited determinants $|\Phi_{i_1 \dots i_n}^{a_1 \dots a_n}\rangle$ with $n \leq m_A$, in an analogous way to the ground-state CC procedure, we can proceed to solve the eigenvalue problem

$$\langle\Phi_{i_1 \dots i_n}^{a_1 \dots a_n}| \left(\bar{H}_{\text{open}}^{(A)} R_{\mu,\text{open}}^{(A)}\right)_C |\Phi\rangle = \omega_\mu^{(A)} r_{\mu,a_1 \dots a_n}^{i_1 \dots i_n}, \quad i_1 < \dots < i_n, \quad a_1 < \dots < a_n, \tag{2.34}$$

where $\omega_\mu^{(A)} = E_\mu^{(A)} - E_0^{(A)}$ are vertical excitation energies, and $\bar{H}_{\text{open}}^{(A)} = \bar{H}^{(A)} - \bar{H}_{\text{closed}}^{(A)} = \bar{H}^{(A)} - E_0^{(A)} \mathbb{1}$ and $R_{\mu,\text{open}}^{(A)} = R_\mu^{(A)} - r_{\mu,0} \mathbb{1}$ are the open parts of $\bar{H}^{(A)}$ and $R_\mu^{(A)}$, which are represented by diagrams having external Fermion lines only. The solution of this eigenvalue problem is usually accomplished iteratively, by exploiting the Davidson's approach [215, 216]. Once $r_{\mu,a_1 \dots a_n}^{i_1 \dots i_n}$ amplitudes are known, the zero-body component of the EOM excitation operator, $R_{\mu,0}^{(A)}$, is calculated *a posteriori* using the following expression:

$$r_{\mu,0}^{(A)} = \frac{\langle\Phi|(\bar{H}_{\text{open}}^{(A)} R_{\mu,\text{open}}^{(A)})_C|\Phi\rangle}{\omega_\mu^{(A)}}. \tag{2.35}$$

2.2 The Active-Space Coupled-Cluster Approaches

The previous section summarizes the conventional SRCC and EOMCC approaches, such as CCSD/EOMCCSD, CCSDT/EOMCCSDT, etc., where the cluster operator, T, and the EOM linear excitation operator, R_μ, are truncated in the many-body expansion at increasing excitation levels (e.g., at T_2 and $R_{\mu,2}$ in CCSD/EOMCCSD, at T_3 and $R_{\mu,3}$ in CCSDT/EOMCCSDT, etc.). Even though these approximations work well in many situations, low-order truncations of CC, such as CCSD/EOMCCSD, are often times not sufficient when studying systems with stronger multi-reference character, such as those resulting from bond breaking events and excited states dominated by two-electron transitions. On the other hand, the higher-level SRCC and EOMCC methods that include

higher-than-two-body components of T and R_μ, and which can handle these more challenging situations, expose us to prohibitively expensive computer costs, which as discussed before realize from the steep CPU time-step scaling of $n_o^3 n_u^5$ for CCSDT/EOM-CCSDT, $n_o^4 n_u^6$ for CCSDTQ/EOMCCSDTQ, etc. In response to this situation, one can employ the active-space CC/EOMCC approaches, which allow us to include higher-than-doubly excited cluster components as in multi-reference approaches, but without turning to the genuine MRCC considerations, while also avoiding the prohibitive costs of the CCSDT/EOMCCSDT, CCSDTQ/EOMCCSDTQ, etc. In the following paragraphs, we will provide an introduction to the theory behind active-space CC and EOMCC methodology, focusing on the popular CCSDt/EOMCCSDt and CCSDtq/EOMCCSDtq approximations.

In the active-space CC and EOMCC approaches, the cluster operator T and the EOM linear excitation operator $R_\mu^{(A)}$ are typically approximated in such a way that T_1, T_2, $R_{\mu,1}$, and $R_{\mu,2}$ are treated fully, but higher-than-two-body components of T and R_μ are selected with the help of active orbitals. In order to accomplish this, the available occupied and unoccupied spin-orbitals, in $|\Phi\rangle$, which define the basis set, are further subdivided into four different groups, that facilitate the selection of the dominant amplitudes in T_3, $R_{\mu,3}$, T_4, $R_{\mu,4}$, etc. Under this scheme, the occupied spin-orbitals are distributed into the core, labeled with lower-case bold indices, $\mathbf{i}, \mathbf{j}, \ldots$, and active occupied, labeled with upper-case bold indices, $\mathbf{I}, \mathbf{J}, \ldots$ classes, while the unoccupied spin-orbitals are divided into the active unoccupied, labeled with upper-case bold indices, $\mathbf{A}, \mathbf{B}, \ldots$, and virtual, labeled lower-case bold indices, $\mathbf{a}, \mathbf{b}, \ldots$ ones. If the active/inactive character of a given index does not have to be specified, we continue using symbols in italics ($i, j, \ldots$ for the occupied and $a, b, \ldots$ for the unoccupied spin-orbitals). The active occupied and unoccupied spin-orbital groups are defined by including valence orbitals, such that the occupied (e.g., HOMO, HOMO-1, HOMO-2, etc.) and unoccupied (LUMO, LUMO+1, LUMO+2, etc.) spin-orbitals relevant to the target system are considered in the active

26
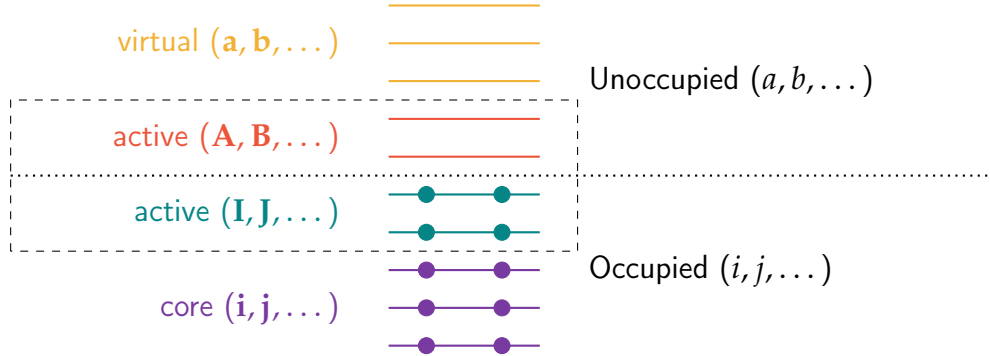
space.

Figure 2.2: An illustration of the orbital classification scheme used in the active-space CC framework. The active occupied and unoccupied orbitals are defined using the pair of numbers, (N_o, N_u), such that N_o is the number orbitals counting from HOMO down (e.g., HOMO, HOMO-1, HOMO-2, etc.) and N_u is the number of orbitals counting from the LUMO up (e.g., LUMO, LUMO+1, LUMO+2, etc.).

Once suitable core, active occupied, active unoccupied, and virtual spin-orbitals are selected for a particular system, we define the cluster and linear-excitation operators entering the active-space CC and EOMCC methods. This is achieved by down-selecting higher-than-two-body components of T and R_μ to within internal and semi-internal excitations that use active spin-orbitals indices in their definitions. For instance, in the CCSDt/EOMCCSDt and CCSDtq/EOMCCSDtq approaches, the relevant operators are defined as follows:

$$T^{(\text{CCSDt})} = T_1 + T_2 + t_3, \qquad t_3 = \sum_{\substack{\mathbf{I}>j>k \\ a>b>\mathbf{C}}} t_{ab\mathbf{C}}^{\mathbf{I}jk} E_{\mathbf{I}jk}^{ab\mathbf{C}}, \tag{2.36}$$

$$T^{(\text{CCSDtq})} = T_1 + T_2 + t_3 + t_4, \quad t_4 = \sum_{\substack{\mathbf{I}>\mathbf{J}>k>l \\ a>b>\mathbf{C}>\mathbf{D}}} t_{ab\mathbf{CD}}^{\mathbf{IJ}kl} E_{\mathbf{IJ}kl}^{ab\mathbf{CD}}, \tag{2.37}$$

$$R_\mu^{(\text{EOMCCSDt})} = r_{\mu,0}\mathbb{1} + R_{\mu,1} + R_{\mu,2} + r_{\mu,3}, \qquad r_{\mu,3} = \sum_{\substack{\mathbf{I}>j>k \\ a>b>\mathbf{C}}} r_{\mu,ab\mathbf{C}}^{\mathbf{I}jk} E_{\mathbf{I}jk}^{ab\mathbf{C}} \tag{2.38}$$

$$R_\mu^{(\text{EOMCCSDtq})} = r_{\mu,0}\mathbb{1} + R_{\mu,1} + R_{\mu,2} + r_{\mu,3} + r_{\mu,4}, \quad r_{\mu,4} = \sum_{\substack{\mathbf{I}>\mathbf{J}>k>l \\ a>b>\mathbf{C}>\mathbf{D}}} r_{\mu,ab\mathbf{CD}}^{\mathbf{IJ}kl} E_{\mathbf{IJ}kl}^{ab\mathbf{CD}},$$

(2.39)

where the $t_3/r_{\mu,3}$, operator is defined by restricting an occupied and unoccupied index to within the active space, and the $t_4/r_{\mu,4}$, by restricting two occupied and unoccupied indices in the same way.

The ground-state CCSDt amplitudes and energy are obtained by solving SRCC equations, similar to those introduced in the previous section. We project the connected cluster form of the Schrödinger equation, $(H e^{T_1+T_2+t_3})_C |\Phi\rangle = E_0 |\Phi\rangle$, on all singly, all doubly, and a subset of the triply excited determinants corresponding to the t_3 operator, which yields the system of equations

$$\langle \Phi_i^a | \bar{H}^{(\text{CCSDt})} | \Phi \rangle = 0, \tag{2.40}$$

$$\langle \Phi_{ij}^{ab} | \bar{H}^{(\text{CCSDt})} | \Phi \rangle = 0, \tag{2.41}$$

and

$$\langle \Phi_{\mathbf{I}jk}^{ab\mathbf{C}} | \bar{H}^{(\text{CCSDt})} | \Phi \rangle = 0, \tag{2.42}$$

where $\bar{H}^{(\text{CCSDt})} = e^{-(T_1+T_2+t_3)} H e^{T_1+T_2+t_3} = (H e^{T_1+T_2+t_3})_C$. As before, once solved, we can use the resulting cluster amplitudes and ground-state energy to diagonalize the similarity transformed Hamiltonian of CCSDt, $\bar{H}^{(\text{CCSDt})}$, in the space spanned by $|\Phi_i^a\rangle$, $|\Phi_{ij}^{ab}\rangle$, and $|\Phi_{\mathbf{I}jk}^{ab\mathbf{C}}\rangle$. The result of this operation are the excited state energies, $E_\mu^{(\text{CCSDt})}$, and linear excitation amplitudes in, $R_\mu^{(\text{EOMCCSDt})}$, which define the EOM-CCSDt approach. In the case of the CCSDtq and EOMCCSDtq approaches, we proceed in a similar manner, extending the definitions of T and R_μ to include the corresponding target excitations.

As one can see from the above description, the CCSDt and EOMCCSDt equations are in essence the full CCSDT and EOMCCSDT equations, but using a small subset of

28

triply excited components in T and R_μ. Similarly, the CCSDtq and EOMCCSDtq equations are the CCSDTQ and EOMCCSDTQ equations using small subsets of triples and quadruples. This has two main advantages. First, these methods allow us to access the benefits of high-level CC approaches in dealing with higher-order correlation effects, that CCSD and EOMCCSD fail to capture entirely, especially when bond breaking events, bi-radicals, and excited electronic states with considerable two-electron excitations are studied. This has been established in many publications [41, 88, 89, 155, 157, 160–162], where the active-space CC and EOMCC calculations are shown to be typically in very good agreement with their full CC and EOMCC parents. And second, the active-space approaches maintain a computational cost similar to the one of CCSD and EOMCCSD methods. Usually, the numbers of active occupied and unoccupied orbitals, N_o and N_u, respectively, are quite small relative to the basis set ($N_o < n_o$ and $N_u \ll n_u$), which means that the CPU time scaling associated with CCSDt/EOMCCSDt and CCSDtq/EOMCCSDtq calculations, and given by $N_o N_u n_o^2 n_u^4$, and $N_o^2 N_u^2 n_o^2 n_u^4$, respectively, corresponds to CCSD CPU time steps times a small prefactor given by the orbitals in the active space.

2.3 The CC(P;Q) Formalism

As mentioned before, the CC(P;Q) hierarchy has the potential to address the key challenges affecting SRCC approaches, especially when describing strongly correlated systems [51, 164, 170, 172, 173, 202, 217]. This includes the huge computational costs associated with the traditional SRCC methods where the many-body expansion of the cluster operator is truncated at higher-than-double excitations, such as the CCSDT/EOM-CCSDT or CCSDTQ/EOMCCSDTQ approaches, the inability of the MMCC based correction δ_μ, featured in methods including the δ-CR-CC(2,3) and δ-CR-EOMCC(2,3), of allowing the lower-order cluster components to relax in the presence of higher-than-

double excitations, and the difficulty, in some situations, of the active-space CC approaches in capturing all of the relevant correlation effects due to a reduced treatment of higher-than-two-body components, as in methods such as CCSDt/EOMCCSDt, CCSDtq/EOMCCSDtq, etc. For example, the CC(P;Q) scheme merges the strengths of the active-space CC methods, which allows the relaxation of lower-order components of T and R_μ in the presence of higher-than-double excitations with the power of the MMCC formalism that provides an efficient way to correct for the missing correlation effects initially disregarded in the active-space considerations

In general, the CC(P;Q) framework provides a systematic way to define iterative CC and EOMCC calculations using any conventional or unconventional truncations of T and R_μ, and to correct their resulting energies, *a posteriori* [51, 163, 164]. The core element behind CC(P;Q), is the definition of two subspaces of the many-electron Hilbert space, $\mathscr{H}^{(P)}$ and $\mathscr{H}^{(Q)}$, called the P and Q spaces, respectively, where $\mathscr{H}^{(Q)} \subseteq (\mathscr{H}^{(0)} \oplus \mathscr{H}^{(P)})^{\perp}$. The P space is used to define the excited cluster and linear excitation components entering CC and EOMCC equations, which are solved iteratively during the first step of the CC(P;Q) calculations. Subsequently, the Q space is utilized to capture the missing correlation effects, described by determinants not included in the P space, by performing a non-iterative correction to the energies $E_\mu^{(P)}$ ($\mu \geq 0$). For example, if the P space is chosen to include all singles and doubles, and a subset of the triples or triples and quadruples, such as CCSDt/EOMCCSDt and CCSDtq/EOMCCSDtq, and the Q space is defined as the space spanned by the remaining triply or triply and quadruply excited determinants, the resulting CC(P;Q) schemes are CC(t;3), CC(t,q;3), CC(t,q;3,4), etc. [51, 164, 170, 172, 173, 217] where the letter symbols explain what higher-than-double excitations are included in the P-space CC/EOMCC calculations, and the integer numbers define the level of the missing excitations included in the Q space, which lead to the corrections $\delta_\mu(P; Q)$. Analogously, if we choose the P space to be spanned by all singly and doubly excited determinants, and we select the Q space

to be spanned by all triply, or triply and quadruply excited determinants, we obtain the CR-CC(2,3) and CR-CC(2,4) methods respectively.

Although these examples, which correspond to CC(P;Q) methods implemented by our group, and which utilize P spaces borrowed from standard and active-space CC approaches, the CC(P;Q) theory is flexible enough to allow for the selection of any conventional or unconventional P and Q spaces, as long as they are proper sub-spaces of the N-particle Hilbert space.

With the intent of emphasizing this, we begin by rewriting the equations defining SRCC hierarchies, Eqs. (2.24) and (2.28), such that

$$|\Psi_0^{(P)}\rangle = e^{T^{(P)}}|\Phi\rangle \tag{2.43}$$

and

$$|\Psi_\mu^{(P)}\rangle = R_\mu^{(P)} e^{T^{(P)}}|\Phi\rangle, \tag{2.44}$$

explicitly specify the restriction of the many-body expansion of $T^{(P)}$ and $R_\mu^{(P)}$ to the P space. In this context, the cluster and linear excitation operators are defined by

$$T^{(P)} = \sum_k t_k E_k, \quad |\Phi_k\rangle \in \mathscr{H}^{(P)} \tag{2.45}$$

and

$$R_\mu^{(P)} = r_{\mu,0}\mathbb{1} + \sum_k r_{\mu,k} E_k, \quad |\Phi_k\rangle \in \mathscr{H}^{(P)}, \tag{2.46}$$

where k is an index defining the determinants that span the P space, t_k and $r_{\mu,k}$ are the amplitudes of the cluster and linear excitation operators, respectively, and E_k is the excitation operator that generates the excited determinants, $|\Phi_k\rangle$, relative to the reference determinant, $|\Phi\rangle$, such that

$$|\Phi_k\rangle = E_k |\Phi\rangle, \quad |\Phi_k\rangle \in \mathscr{H}^{(P)}. \tag{2.47}$$

With these definitions, we proceed once again to project the connected cluster form of the Schrödinger equation, $\bar{H}^{(P)}|\Phi\rangle = E_0^{(P)}|\Phi\rangle$, utilizing the ansätze (2.43) and (2.44),

but this time onto the excited determinants, $|\Phi_k\rangle$, included in the P space. This yields the following set of equations:

$$\langle \Phi_k | \bar{H}^{(P)} | \Phi \rangle = 0, \quad |\Phi_k\rangle \in \mathscr{H}^{(P)}, \tag{2.48}$$

where $\bar{H}^{(P)} = e^{-T^{(P)}} H e^{T^{(P)}} = (He^{T^{(P)}})_C$ is the similarity transformed Hamiltonian corresponding to the P space. We continue solving these equations in the usual way, to obtain the cluster amplitudes entering $T^{(P)}$ and the associated ground-state energy $E_0^{(P)}$.

Using the EOM framework, the amplitudes defining the linear excitation operator $R_\mu^{(P)}$ and the corresponding excited-state energies $E_\mu^{(P)}$ are obtained by diagonalizing the similarity transformed Hamiltonian $\bar{H}^{(P)}$ in the P space, resulting in the system of equations

$$\langle \Phi_k | \left(\bar{H}_{\text{open}}^{(P)} R_{\mu,\text{open}}^{(P)} \right)_C | \Phi \rangle = \omega_\mu^{(P)} r_{\mu,k}, \quad |\Phi_k\rangle \in \mathscr{H}^{(P)}. \tag{2.49}$$

And finally, the zero-body excitation operator, $R_{\mu,0}^{(P)}$, is calculated using a similar expression to Eq. (2.35) given by

$$r_{\mu,0}^{(P)} = \frac{\langle \Phi | (\bar{H}_{\text{open}}^{(P)} R_{\mu,\text{open}}^{(P)})_C | \Phi \rangle}{\omega_\mu^{(P)}}. \tag{2.50}$$

Once all cluster and linear excitation amplitudes and the corresponding energies are computed, we employ a treatment based on the MMCC theory to compute, *a posteriori*, state-specific, and non-iterative corrections $\delta_\mu(P; Q)$, which recover mostly dynamical correlation effects contained in the Q space. In order to derive the formulation of the correction, $\delta_\mu(P; Q)$, let us begin with the asymmetric energy expression,

$$E_\mu = \frac{\langle \Psi_\mu | H R_\mu^{(P)} e^{T^{(P)}} | \Phi \rangle}{\langle \Psi_\mu | R_\mu^{(P)} e^{T^{(P)}} | \Phi \rangle}. \tag{2.51}$$

In this expression, $\langle \Psi_\mu |$ is the FCI bra state for the ground ($\mu = 0$) or excited ($\mu > 0$) states, and E_μ is the corresponding energy, which is exact regardless of the selection of the P space. This fact can be easily proven by applying the turn-over rule and acknowledging that, $\langle \Psi_\mu | H = \langle \Psi_\mu | E_\mu$, since we are assuming the $\langle \Psi_\mu |$ is the exact, FCI, state.

The biorthogonal formulation of the MMCC theory, which is the foundation beneath the CR-CC methods, such as CR-CC(2,3), CR-CC(2,4), etc., is characterized by the reparametrization of the FCI bra state, given by

$$\langle \Psi_\mu | = \langle \Phi | \mathscr{L}_\mu e^{-T^{(P)}}, \tag{2.52}$$

where $\mathscr{L}_\mu$ is a linear de-excitation operator defined by a many-body expansion of elementary particle-hole de-excitation operators,

$$\mathscr{L}_\mu = \delta_{\mu,0} \mathbb{1} + \sum_k \ell_{\mu,k} (E_k)^\dagger, \quad \langle \Phi_k | \in \mathscr{H}, \tag{2.53}$$

where the determinants, $\langle \Phi_k | = \langle \Phi | (E_k)^\dagger$, span the complete N-electron Hilbert space $\mathscr{H}$. Clearly, the operator $\mathscr{L}$ is always defined, since one can always transform the expression of Eq. (2.52), into the left-eigenvalue problem

$$\langle \Phi | \mathscr{L}_\mu \bar{H}^{(P)} = E_\mu \langle \Phi | \mathscr{L}_\mu, \tag{2.54}$$

which resembles the connected cluster form of the Schrödinger equation and which is fully equivalent to $\langle \Psi_\mu | H = E_\mu \langle \Psi_\mu |$. By replacing $\langle \Psi_\mu |$ in the asymmetric energy expression given above, Eq. (2.52), and by inserting the resolution of identity in the N-electron Hilbert space, $\mathbb{1} = |\Phi\rangle \langle \Phi | + \sum_k |\Phi_k\rangle \langle \Phi_k |$, between $\mathscr{L}_\mu$ and $e^{-T^{(P)}}$, we obtain the following energy expression:

$$E_\mu = \frac{\langle \Phi | \mathscr{L}_\mu [|\Phi\rangle \langle \Phi | + \sum_k |\Phi_k\rangle \langle \Phi_k |] \bar{H}^{(P)} R_\mu^{(P)} |\Phi\rangle}{\langle \Phi | \mathscr{L}_\mu R_\mu^{(P)} |\Phi\rangle}, \tag{2.55}$$

which can be further simplified by imposing the normalization condition, $\langle \Phi | \mathscr{L}_\mu R_\mu^{(P)} |\Phi\rangle = 1$, and rearranging terms such that,

$$E_\mu = E_\mu^{(P)} + \sum_k \ell_{\mu,k} \langle \Phi_k | \bar{H}^{(P)} R_\mu^{(P)} |\Phi\rangle, \quad |\Phi_k\rangle \notin \mathscr{H}^{(P)}. \tag{2.56}$$

We name the quantities $\mathfrak{M}_{\mu,k}(P) = \langle \Phi_k | \bar{H}^{(P)} R_\mu^{(P)} |\Phi\rangle$ moments of CC/EOMCC equations. It is worth noting that the values of $\mathfrak{M}_{\mu,k}$ consist of the usual connected cluster

form of the Schrödinger equation, but projected onto excited determinants spanning the orthogonal space to the P space, unlike the usual procedure for CC/EOMCC equations in which $\bar{H}^{(P)}R_\mu^{(P)}|\Phi\rangle$ is projected only onto excited determinants spanning the P space.

Up until this point, we have operated with the assumption that $\langle\Psi_\mu|$ is the exact, FCI, wave function. However, since the purpose of developing a method based on the MMCC theory is to avoid the FCI problem itself, we must find an alternative approximation. In the CC(P;Q) framework, we do this by rewriting the operator $\mathscr{L}_\mu$, such that it is approximated by

$$\mathscr{L}_\mu \approx L_\mu^{(P)} + \mathscr{L}_\mu^{(Q)}(P), \tag{2.57}$$

where the linear de-excitation operator, $L_\mu^{(P)}$, defined by

$$L_\mu^{(P)} = \delta_{\mu,0}\mathbb{1} + \sum_{|\Phi_k\rangle\in\mathscr{H}^{(P)}} l_{\mu,k}(E_k)^\dagger. \tag{2.58}$$

belongs to the CC/EOMCC left or bra state,

$$\langle\tilde{\Psi}_\mu^{(P)}| = \langle\Phi|L_\mu^{(P)}e^{-T^{(P)}}, \tag{2.59}$$

and satisfies the biorthonormality condition:

$$\langle\Phi|L_\mu^{(P)}R_\nu^{(P)}|\Phi\rangle = \delta_{\mu\nu}. \tag{2.60}$$

The amplitudes $l_{\mu,k}$, entering $L_\mu^{(P)}$, are obtained by solving the left CC eigenvalue problem

$$\delta_{\mu,0}\langle\Phi|\bar{H}_{\text{open}}^{(P)}|\Phi_k\rangle + \langle\Phi|L_{\mu,\text{open}}^{(P)}\bar{H}_{\text{open}}^{(P)}|\Phi_k\rangle = \omega_\mu^{(P)}l_{\mu,k}, \quad |\Phi_k\rangle\in\mathscr{H}^{(P)}. \tag{2.61}$$

Since the similarity transformed Hamiltonian, $\bar{H}_{\text{open}}^{(P)}$, and the vertical excitation energies, $\omega_\mu^{(P)}$, are already known from the previous CC and EOMCC steps, one can solve for the amplitudes $l_{\mu,k}$, by treating the left eigenvalue problem as a system of homogeneous equations.

The remaining operator $\mathscr{L}_\mu^{(Q)}(P)$ in Eq. (2.57) approximates the many-body components of the operator $\mathscr{L}_\mu$ spanning the Q space. The amplitudes, $\ell_{\mu,k}$, entering this operator satisfy the system of equations:

$$\langle\Phi|L_\mu^{(P)}\bar{H}^{(P)}|\Phi_k\rangle + \langle\Phi_{k'}|\bar{H}^{(P)}|\Phi_k\rangle\,\ell_{\mu,k'} = E_\mu\ell_{\mu,k}, \quad \langle\Phi_{k'}|,|\Phi_k\rangle \in \mathscr{H}^{(Q)}, \qquad (2.62)$$

which we can solve approximately by letting $E_\mu \approx E_\mu^{(P)}$ and by neglecting terms in $\langle\Phi_{k'}|\bar{H}^{(P)}|\Phi_k\rangle$, such that the values of $D_{\mu,k}$ entering

$$\ell_{\mu,k} = \frac{\langle\Phi|L_\mu^{(P)}\bar{H}^{(P)}|\Phi_k\rangle}{D_{\mu,k}}, \qquad |\Phi_k\rangle \in \mathscr{H}^{(Q)}, \qquad (2.63)$$

can be systematically approximated by Epstein-Nesbet-like (EN) denominators,

$$D_{\mu,k}^{\mathrm{EN}} = E_\mu^{(P)} - \langle\Phi_k|\bar{H}^{(P)}|\Phi_k\rangle, \qquad (2.64)$$

Møllet–Plesset-like (MP) denominators

$$D_{\mu,k}^{\mathrm{MP}} = E_\mu^{(P)} - \langle\Phi_k|H|\Phi_k\rangle, \qquad (2.65)$$

or other schemes. Since using EN denominators leads to considerably more accurate energy results, in the remaining chapters of this thesis, we will focus our discussion on CC(P;Q) energies calculated using EN denominators.

Now, using Eq. (2.56), we can write the central equation of CC(P;Q), namely

$$E_\mu^{(P+Q)} \equiv E_\mu^{(P)} + \delta_\mu(P;Q), \qquad (2.66)$$

where, $\delta_\mu(P;Q)$ is the non-iterative correction to the ground- ($\mu = 0$) and excited-state energies ($\mu > 0$) computed in the P space, $E_\mu^{(P)}$, due to the missing correlation effects described by the determinants spanning the Q space, and which is defined as

$$\delta_\mu(P;Q) = \sum_{\substack{|\Phi_k\rangle\in\mathscr{H}^{(Q)} \\ \mathrm{rank}(|\Phi_k\rangle)\leq\min(N_\mu^{(P)},\Xi(Q))}} \ell_{\mu,k}(P)\,\mathfrak{M}_{\mu,k}(P), \qquad (2.67)$$

35

where $N_\mu^{(P)}$ designates the highest many-body rank of the excited determinants $|\Phi_k\rangle$ for which the moment $\mathfrak{M}_{\mu,k}(P)$ is non-zero, and $\Xi^{(Q)}$ is the highest many-body rank of $|\Phi_k\rangle$ in $\mathscr{H}^{(Q)}$.

As it has been said before, the non-iterative step of the CC(P;Q) formalism is a generalization of the corrections used in the CR-CC and CR-EOMCC approaches based on the MMCC equations, that extends the notion of rank-incremental many-body expansions, as employed in standard CC methods, to account for any conventional or unconventional definition of the underlying P and Q spaces. With this in mind, one can think of exploiting other ways of defining P and Q spaces, which do not rely on manual intervention, such as in the non-black-box active-space CC approaches [163, 164, 218]. In the remaining sections of this chapter, we will introduce the theories behind the automated P space selection, proposed in this Ph.D. thesis, and that rely on the stochastic sampling of the Hilbert space, as provided by the i-CIQMC approaches.

2.4   The FCIQMC and its Initiator Approximation (i-FCIQMC)

FCIQMC is a new quantum Monte Carlo methodology [1, 189–197, 219–227] which takes its foundations from the well-known DMC methods, by exploiting the propagation of the imaginary-time Schrödinger equation to formulate diffusion like equations in the space of Slater determinants. However, in contrast with DMC, which relies on long numerical integrations in the $3N$-electron coordinate space, FCIQMC uses a population dynamics algorithm that employs a set of walkers designed to sample the FCI space. These walkers are discrete counting variables, which also carry a sign, and inhabit various determinants of the FCI space, evolving through imaginary-time according to a simple set of rules, which allows them to converge onto the exact, FCI, wave function in the limit of infinite imaginary time.

To explain the FCIQMC approach in detail, we must first discuss the foundations of

the deterministic FCI approach. As stated in the Introduction, FCI is a method based on the fact that the numerically exact solution of the Schrödinger equation can be obtained by considering all possible electron-electron interactions via the diagonalization of the Hamiltonian matrix, $\mathbf{H}$. Doing so, we obtain the exact, FCI, energies, $E_\mu^{(\text{FCI})}$ and corresponding wave functions, $|\Psi_\mu^{(\text{FCI})}\rangle$ ($\mu \geq 0$), which exist as linear combinations of the reference and all excited Slater determinants, such that $c_{\mu,k}$ are the coefficients constituting the vector $\mathbf{C}_\mu$. The following expressions define the time-independent matrix eigenvalue problem and FCI wave function, that we will use throughout this work:

$$\mathbf{H}\mathbf{C}_\mu = E_\mu^{(\text{FCI})}\mathbf{C}_\mu, \tag{2.68}$$

$$|\Psi_\mu^{(\text{FCI})}\rangle = \sum_k c_{\mu,k} |\Phi_k\rangle . \tag{2.69}$$

The time-dependent Schrödinger equation is given by

$$\frac{\partial |\Psi_\mu\rangle}{\partial \tau} = -H |\Psi_\mu\rangle , \tag{2.70}$$

where τ is the imaginary time, also known as the Wick's rotation of time, and H is the usual electronic Hamiltonian comprising one- and two-body terms. The time-dependent Schrödinger equation can be integrated with respect to τ, which results in the imaginary time propagation, $e^{-\tau H} |\Psi_\mu\rangle$. The propagator, $e^{-\tau H}$, provides a natural way to evolve a trial wave function, so that the exact, FCI, ground-state wave function projects out of the infinite imaginary-time limit, given by

$$|\Psi_0^{(\text{FCI})}\rangle \propto \lim_{\tau \to \infty} e^{-\tau(H-E_0)} |\tilde{\Psi}\rangle , \qquad \langle\tilde{\Psi}|\Psi_0\rangle \neq 0, \tag{2.71}$$

where E_0 is the exact ground-state energy, required to prevent the above expression from vanishing at the infinite time limit, and $|\tilde{\Psi}\rangle$ is an arbitrary trial wave function, whose only required condition is to possess a non-zero overlap with the target state $|\Psi_0^{(\text{FCI})}\rangle$. One can think of this process using the language of spectroscopy by imagining a superposition state produced by a photon absorption (represented by $|\tilde{\Psi}\rangle$), which decays with time, eventually reaching the ground state (represented by $|\Psi_0^{(\text{FCI})}\rangle$).

By writing the ground-state FCI wave function, in which the time-independent CI coefficients $c_{0,k}$ are replaced with their time dependent $c_{0,k}(\tau)$ counterparts, such that

$$|\Psi_0(\tau)\rangle = c_{0,0}(\tau)|\Phi\rangle + \sum_k c_{0,k}(\tau)|\Phi_k\rangle, \tag{2.72}$$

and by inserting this expression into Eq. (2.70), we obtain a system of diffusion-like equations

$$-\frac{\partial c_{0,k}(\tau)}{\partial \tau} = (H_{kk} - S)c_{0,k}(\tau) + \sum_{l(\neq k)} H_{kl}c_{0,l}(\tau). \tag{2.73}$$

If $S \to E_0$, $\lim_{\tau \to \infty} = \frac{\partial c_{0,k}(\tau)}{\partial \tau}$, then $\sum_l H_{kl}c_{0,l}(\infty) = E_0 c_{0,k}(\infty)$, which is equivalent to recovering the time-independent matrix eigenvalue problem of Eq. (2.68).

After a deeper examination of these equations, with the goal of developing an algorithm suitable for Monte Carlo simulations, one can think of them as a two-step update process of the CI coefficients, such that

$$c_{0,k} - \partial\tau(H_{kk} - E_0)c_{0,k} \to c_{0,k}, \tag{2.74}$$

and

$$c_{0,k} - \partial\tau \sum_l H_{lk}c_{0,l} \to c_{0,k}, \tag{2.75}$$

at every iteration over imaginary time. These expression map naturally to the population dynamics algorithm which constitutes the backbone of the FCIQMC methodology. In it, the CI coefficients are approximated using discrete random walkers, labeled with indices α, and having a sign $s_\alpha = \pm 1$, which are allowed to inhabit one Slater determinant, $|\Phi_k\rangle$, in such a way that the coefficients c_k[1] become proportional to the amount of walkers residing in a given determinant

$$c_k \propto N_k = \sum_\alpha s_\alpha \delta_{k,k_\alpha}. \tag{2.76}$$

---

[1]Since we will describe the ground-state FCIQMC method in this section, we are assuming all indices k and l to belong to the ground state (e.g., $c_k = c_{0,k}$, $N_k = N_{0,k}$, etc.)

The total walker population in a FCIQMC calculation, N_w, is obtained by summing the absolute value of the walker population inhabiting each determinant, $|\Phi_k\rangle$, such that

$$N_w = \sum_k |N_k|. \tag{2.77}$$

The population dynamics algorithm consists in the following three steps derived from Eq. (2.73):

1. **The spawning step:** For each of the available walkers, α, obtained from previous time steps in the simulation, the spawning of a new walker is attempted on a determinant $|\Phi_l\rangle$, selected at random with probability p_{gen}, from the set of determinants connected through the Hamiltonian to the one, $|\Phi_{k_\alpha}\rangle$, populated by the original walker attempting the spawning event. In this process, the probability of a successful spawning event is given by the equation

$$p_s(l|k_\alpha) = \frac{\tau|H_{lk_\alpha}|}{p_{\text{gen}}(l|k_\alpha)}, \tag{2.78}$$

which is proportional to the off-diagonal elements of H. The spawning step can be thought of as the step through which new determinants are discovered and is responsible for the initial increase of the walker population.

2. **The diagonal death/cloning step:** Each of the walkers in the list is cloned or removed from the simulation based on the diagonal elements of the Hamiltonian using the probability

$$p_d(k_\alpha) = \tau(H_{k_\alpha k_\alpha} - S), \tag{2.79}$$

where S is an energy value, known as the shift energy, that functions as a population modulator (see Eq. (2.80)), and which, in the limit of infinite imaginary time converges toward the ground-state energy E_0. If the value of this probability is positive, the walker is eliminated from the list. Otherwise it is cloned, thereby increasing the population in the determinant inhabited by the original walker. This

step is crucial in determining the importance of the determinants fund during the *spawning step*.

3. **The annihilation step:** In this third and final step, all the available walkers are inspected with the goal of removing pairs that populate the same determinant but possess different signs. This is done to stabilize the sign of the c_k coefficients and is a crucial for the convergence of FCIQMC calculations.

To manage the total population of walkers during the calculation, attention must be paid to the aforementioned shift energy. Let us recall that the infinite time limit expression, given in Eq. (2.71), requires the ground-state energy to prevent the limit from vanishing. Because this energy is unknown at the beginning of any FCIQMC calculation, it is necessary to replace E_0 with a variable, S. With this substitution, the infinite-time limit can fall into the following cases:

$$\lim_{\tau \to \infty} e^{(H-S)\tau} |\tilde{\Psi}\rangle = \begin{cases} c_0 |\Psi_0\rangle & \text{for } S = E_0 \\ \infty & \text{for } S > E_0 \\ 0 & \text{for } S < E_0 \end{cases} \quad (2.80)$$

Since the distribution of the total walker population, N_w, over the Slater determinants spanning the FCI space, and defining the FCIQMC wave function at every time step, changes dynamically during the propagation, we can use their variation in time as a cue to determine the value of S. For example, if N_w increases over a period of imaginary-time, we can stabilize it by lowering the value of S. On the other hand, if the N_w is falling, we can counteract this by increasing the value of S. Therefore, by adjusting the value of the shift energy, such that the total population is kept stable in time, one can estimate the ground-state energy E_0, since $S = E_0$ at the limit of infinite imaginary time.

The simulation algorithm can be now summarized. An initial amount of walkers is allocated in a reference determinant of choice, usually the Hartree–Fock determinant.
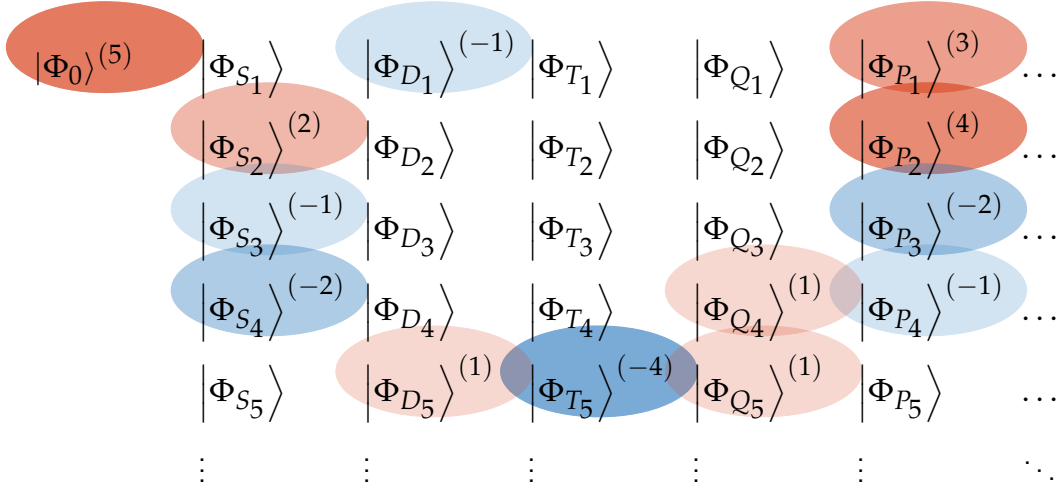
Figure 2.3: An illustration of the population dynamics algorithm in the space of Slater determinants. The red and blue colors denote positive and negative signs, respectively, and their intensity depicts the amount of walkers residing at the determinant enclosed by the ovals. The symbols S_n, D_n, T_n, etc. refer to the singly, doubly, triply, and so forth, excited determinants.

Next, the population dynamics algorithm is allowed to operate in a loop repeating for each time step $\Delta\tau$, while the shift energy is kept above the expected E_0 value (this is usually done by keeping $S = E_0^{(\mathrm{HF})}$, since the HF energy will always lie above the FCI energy due to the variational principle). During the initial stages of the simulation, the setup $S > E_0$ will promote the growth of the walker population, N_w, until a plateau is reached, which is system-dependent and defines the critical walker population N_c. After some time, and while keeping $S > E_0$, the walker population will resume growing, marking the convergence of the sign structure. Following this event, the shift energy is allowed to change in order to maintain the walker population constant for the remaining duration of the simulation.

Since QMC methods are based on Markov chains, the energy resulting from running such simulations is estimated by collecting enough statistically relevant information. FCIQMC methods benefit from two sources of energy data, namely, the shift energies described above, and also the so-called projective enegy, which is computed periodically

41

using the following expression:

$$E(\tau) = \frac{\langle \Phi | H e^{-\tau H} | \Phi \rangle}{\langle \Phi | e^{-\tau H} | \Phi \rangle} = E_0^{\mathrm{HF}} + \sum_l H_{l0} \frac{c_l}{c_0} \approx E_0^{\mathrm{HF}} + \sum_l H_{l0} \frac{N_l}{N_0}. \tag{2.81}$$

Once sufficient information is collected using the shift and projective energy estimators, a so-called reblocking analysis [228] of the data is performed, which removes the time-dependent serial correlation associated with the simulation. The result of the reblocking analysis is considered the FCIQMC energy.

The FCIQMC methods works well, converging onto the FCI energy and wave function with numbers of walkers well below the size of the deterministic FCI space, N_{det}. In some situations, the fraction $f_c = N_c/N_{\mathrm{det}}$, where N_c is the minimum number of walkers required for convergence, can be as low as $f_c \approx 10^{-5}$ [189]. Nevertheless, many other systems require substantially larger fractions of the FCI space to achieve convergence, sometimes nearing $f_c = 1$. This, of course, represents a big problem, since at these higher fractions, the computational costs of FCI and FCIQMC become very similar.

In response to this issue, the initiator approximation was introduced [1]. This improvement over FCIQMC consists of a set of simple rules that accelerate the convergence of FCIQMC by realizing that the key source of noise, leading to the high f_c requirements, was the inability of FCIQMC to converge on a correct sign structure of the wave function. To ameliorate this, Professor Alavi and co-workers prescribed a modification of the rules for the spawning of new walkers, such that a new spawned walker would remain in the simulation only if (i) it was created by an *initiator* determinant, (ii) was created on a determinant that already had a walker population, and (iii) was spawned more than once during the same iteration on a previously unoccupied determinant, with the same sign, and by two different parent determinants (see Figure 2.4 for a pictorial representation). All of these events are deemed to be sign-coherent and, thus, preserve the sign structure of the simulation, considerably reducing the noise and improving the convergence of the *i*-FCIQMC simulations. The way determinants obtain the initiator status is by collecting a sufficiently large population, n_a, which is commonly kept at
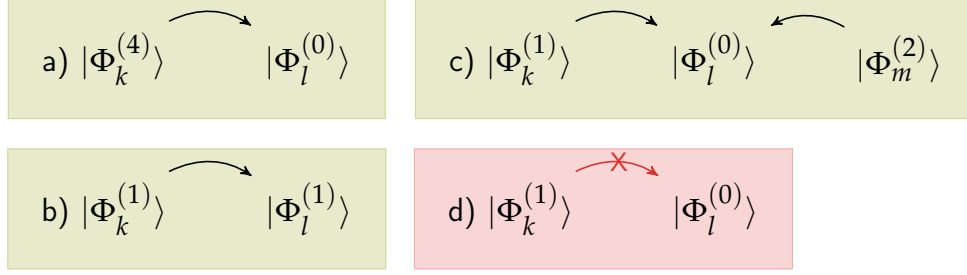
Figure 2.4: Depiction of the allowed and disallowed spawning events in the initiator FCIQMC approach. The number in parenthesis represents the walker population at a given determinant. In this example, the initiatior determinant threshold is set to the default value of $n_a = 3$. Panel a) shows a spawning event taking place, which is allowed since the population of the parent determinant $|\Phi_k\rangle$ is larger than the threshold $n_a = 3$. Panel b) illustrated the situation where a non-initiator determinant (a determinant whose population is smaller than 3) is allowed to spawn progeny since $|\Phi_l\rangle$ is already populated. Panel c) demonstrates a less probable event, where more than one non-initiator parent determinant spawns on a particular determinant $|\Phi_l\rangle$. If the sign of the spawned walker is the same in both events, it is kept in the simulation. Finally, panel d) shows the case when a spawned walker is rejected due the parent determinant not satisfying the requirements for initiator status.

$n_a = 3$ for most simulations. The advantage of employing the initiator algorithm can be noted in examples such as N_2/cc-pVDZ, in which the i-FCIQMC method reduces the total walker population requirement from 2.7×10^8 to 10^5, while maintaining the accuracy below $1\,mE_h$. Although i-FCIQMC is a great improvement over FCIQMC, it introduces a bias in the QMC algorithm which can result in difficulties in convergence due to systematic undersampling of noninitiator determinants. This issue has been recently addressed by utilizing the energy shift to correct for the undersampling bias [190].

2.5 Merging the FCIQMC Ideas with the CC(*P*;*Q*) Formalism: The Semi-Stochastic CC(*P*;*Q*) Approach

Throughout this document it has been stated that, although the CC(*P*;*Q*) hierarchy of SRCC methods have the potential to address the issue of describing both dynamical and nondynamical correlation effects at very reasonable computational costs, there are still open questions regarding how to select an unconventional *P* space allowing for the

relaxation of lower-order cluster components and yet small enough to reduce the huge costs associated with traditional higher-order methods, such as CCSDT, CCSDTQ, etc.

One of the central goals of this thesis is to answer precisely this question. The proposed strategy consists in developing an automated approach exploiting the ability of the FCIQMC method of providing a compact wave function description by sampling the entire space of Slater determinants. Since FCIQMC simulations are capable of quickly identifying the leading excitations dominating a wave function of a particular electronic state, we can let lists of determinants collected during FCIQMC simulations define the appropriate P space in deterministic CC(P;Q) calculations, which otherwise would have to be defined manually, using some *a priori* knowledge of the wave function.

Before we continue with our explanation on how we merge the FCIQMC methodology with the framework provided by the CC(P;Q) theory, we wish to make one more remark about QMC methods. One interesting feature of the FCIQMC methodology that is not commonly discussed, is the fact that it allows for the computation of truncated CIQMC variants. What this means is that FCIQMC simulation can be restricted to a subspace of all Slater determinants, such that spawning events attempting to create new walkers on determinants of an excitation rank larger than a threshold are automatically rejected. In this way, one can recover truncated CI energetics, such as CISD, CISDT, CISDTQ, etc. out of the same population dynamics algorithm described above. It is also worth mentioning that we can replace the linear wave function ansatz of FCIQMC, with the exponential one characterizing CC, to obtain the CCMC approach, where we replace the walkers of FCIQMC with analogous objects, called excips, which sample the various cluster amplitudes entering the CC wave function ansatz (cf. Refs. [198–202] for further details).

Clearly, the appeal of running such kind of simulation is rather low, since *i*-FCIQMC simulations are already much more tractable than FCI, leading to very high accuracies, which avoid the pitfalls associated with truncated CI methods, such as the size-

extensivity issues and large errors in the correlation energy. However, our interest in such computations lies, first, in the fact that their computer requirements are much lower relative to FCIQMC, and second, in the realization that we are only interested in sampling the leading components defining a subspace of the Hilbert space and not in numerical values. Thus, the quality of the CI coefficients representing the i-CISDT-MC, i-CISDTQ-MC, etc. wave functions does not affect our considerations as will become evident during our discussions in Section 3.1.

Even though many different possibilities can be explored in order to merge the deterministic CC(P;Q) framework with the stochastic CIQMC (or CCMC) methodology, in this work we do this, for both ground- and excited-electronic state calculations, by adopting the following procedure:

1. We begin a CIQMC (or CCMC) simulation, which can be of the truncated CI kind such as i-CISDT-MC, or i-CISDTQ-MC, or i-FCIQMC, for the ground state, by placing a certain amount of walkers on the reference determinant. If targeting excited electronic states in addition to the ground state, and if the target many-electron state has spin and spatial symmetries, we start analogous propagations for the lowest state belonging to the same irreducible representation as each of the excited states of interest.

2. At suitably chosen snapshots along the imaginary-time axis, we extract the lists of determinants holding at least one walker from the simulation (or simulations) started in the previous step. With them, we define the appropriate P spaces in the following manner: For each of the irreducible representations of interest, we define the P space as all singles, all doubles, and a subset of the many-body components of rank $n > 2$, such that their corresponding excited Slater determinants are included in the CIQMC list.

3. We solve the ground-state CC(P) equations using the stochastically determined

P spaces, such that the operator, T, defined in the P space, is given by $T^{(P)} = T_1 + T_2 + T_3^{(MC)} + T_4^{(MC)} + \ldots$. In the solution procedure, we employ the usual iterative approach, characteristic of standard CC methods.

4. Using the similarity transformed Hamiltonian, $\bar{H}^{(P)} = (He^{T^{(P)}})_C$, resulting from the calculation in the previous step, diagonalize $\bar{H}^{(P)}$ in the P spaces resulting from the propagations of the lowest energy states of the irreducible representations corresponding to the excited state of interest, such that the linear excitation operator is defined by $R_\mu^{(P)} = r_{\mu,0}\mathbb{1} + R_{\mu,1} + R_{\mu,2} + R_{\mu,3}^{(MC)} + R_{\mu,4}^{(MC)} + \ldots$.

5. If the target calculation includes the CC(P;Q) correction, once the ground- and excited-state energies are determined in the previous steps, the Q space is defined by including all excited Slater determinants, not captured by CIQMC, and, thus, missing from the P space, up to a given excitation rank. Then, after solving for the amplitudes entering the de-excitation operator $L_\mu^{(P)} = \delta_{\mu,0}\mathbb{1} + L_{\mu,1} + L_{\mu,2} + L_{\mu,3}^{(MC)} + \ldots$, one can compute the amplitudes $\ell_{\mu,k}$ and moments $\mathfrak{M}_{\mu,k}$, in the Q space, such that we obtain the corrections $\delta_\mu(P;Q) = \sum_k \ell_{\mu,k} \langle \Phi_k | \bar{H}^{(P)} | \Phi \rangle$, where $\langle \Phi_k |$ are excited determinants belonging to the Q space.

6. At this point we check for convergence. If the CC(P) or CC(P;Q) energies obtained from the current snapshot of the CIQMC simulation differ by more than a predefined threshold from the previous calculated values, return to the item 2) and continue iterating over the procedure. Otherwise, stop all CIQMC simulations and accept the CC(P) or CC(P;Q) energy as the converged value.
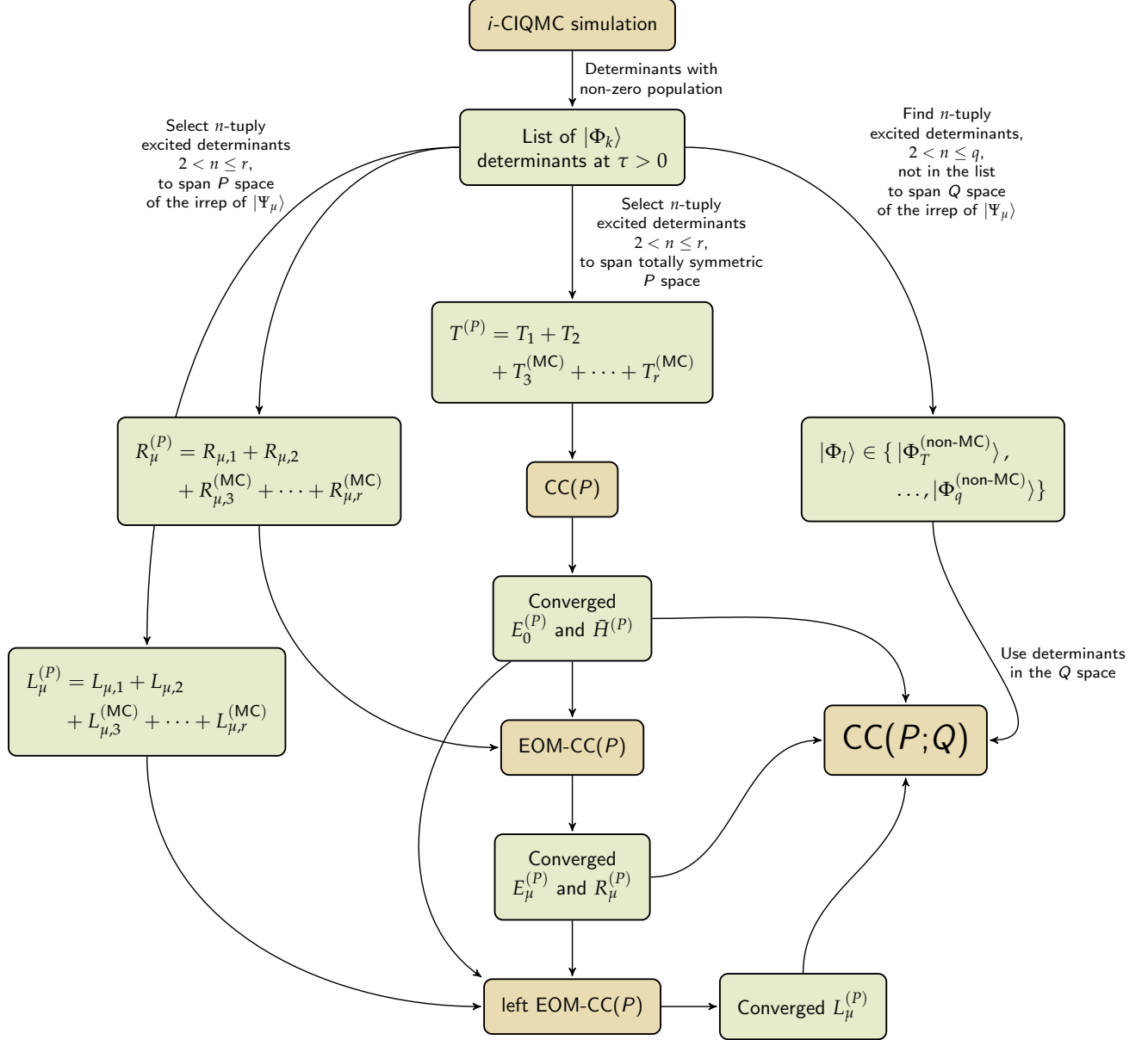
Figure 2.5: Diagram depicting the flow of a semi-stochastic CC(P;Q) calculation using CIQMC. The brown nodes denote QC methods, while the green ones indicate stochastically determined many-body objects shared between the various intermediate computations. As described in the body of this section, the P spaces are defined by all singles, all doubles, and a subset of the higher-than-doubly excited Slater determinants, determined during the CIQMC simulation, up to a maximum many-body rank r. On the other hand, the Q spaces are defined by selecting higher-than-doubly excited Slater determinants, not captured during the corresponding CIQMC simulations, up to a maximum many-body rank q. In the example calculations shown in Chapter 3, the values of r and q were set to 3 when recovering CCSDT energetics, and to $r = 4$ and $q = 3$ when reproducing the higher-level CCSDTQ energies.

In the present work, we employ this procedure using i-CISDT-MC, i-CISDTQ-MC, and i-FCIQMC simulations and define our P spaces using only up to quadruply excited components captured during the stochastic simulations. When correcting for the correlation effects missing from these considerations, we utilized CC(P;Q) energy corrections computed on Q spaces including only triply excited determinants determinants. However, the algorithm presented in this section is applicable to any order truncated i-CIQMC and for any definition of P and Q spaces, including excitation ranks larger than 3 or 4.

As will be shown during the presentation of our results, the key advantage of the merger between the stochastic CIQMC methodology and deterministic CC(P;Q) framework, is that it provides a straightforward solution addressing the non-black-box issue associated with the manual selection of P spaces, such as in the active-space partitioning of the CC(t;3), CC(t,q;3), and others. With our new approach, we can reproduce the energies of the hugely expensive CC methods including higher-than-pair clusters, by solving CC equations and computing moment energy expansions in much smaller subspaces of the Hilbert space. To understand the CPU time-step speedups resulting from the semi-stochastic CC(P;Q) methodology, let us look into the example where CCSDT energies are reproduced via CC(P) calculations where $T^{(P)} = T_1 + T_2 + T_3^{(\mathrm{MC})}$. If we examine the most expensive term given by $\langle \Phi_{ijk}^{abc} | [H, T_3] | \Phi \rangle$ and we let D be the total number of triples entering CCSDT while d is the number of triples captured during the stochastic i-CIQMC simulation, the CPU time-step speedup is proportional to $(\frac{d}{D})^2$.

2.6   The CAD-FCIQMC Method

Another option for the exploitation of stochastic wave functions within the CC framework is to forego the standard and non-standard truncation-based CC approximations described in the first few sections of this chapter and tackle the electronic correlation

problem by directly approaching exact, FCI, energetics. This can be achived by adapting the historical externally corrected CC methods, pioneered more than two decades ago, in order to benefit from the advantages of FCIQMC. The externally corrected CC methods originate from a very insightful observation in CC equations derived from electronic Hamiltonians with up to two-body interactions. If the connected cluster form of the Schrödinger equation, corresponding to the exact single-reference CC theory equivalent to FCI, is projected onto singly and doubly excited determinants, one obtains

$$\langle\Phi_i^a|\,[H_N(\mathbb{1} + T_1 + T_2 + \tfrac{1}{2}T_1^2 + \boldsymbol{T_3} + T_1 T_2 + \tfrac{1}{6}T_1^3)]_C\,|\Phi\rangle = 0, \tag{2.82}$$

$$\begin{aligned}\langle\Phi_{ij}^{ab}|\,[H_N(\mathbb{1} + T_1 + T_2 + \tfrac{1}{2}T_1^2 + \boldsymbol{T_3} + T_1 T_2 + \tfrac{1}{6}T_1^3 \\ + \boldsymbol{T_4} + T_1\boldsymbol{T_3} + \tfrac{1}{2}T_2^2 + \tfrac{1}{2}T_1^2 T_2 + \tfrac{1}{24}T_1^4)]_C\,|\Phi\rangle = 0.\end{aligned} \tag{2.83}$$

Since the CC energy, given by

$$E_0 = \langle\Phi|H|\Phi\rangle + \langle\Phi|[H_N(T_1 + T_2 + \tfrac{1}{2}T_1^2)]_C)|\Phi\rangle, \tag{2.84}$$

only depends on T_1 and T_2, and T_1 and T_2 can be obtained by solving the above equations in the presence of $\boldsymbol{T_3}$ and $\boldsymbol{T_4}$ and no higher order cluster component, one can acquire the exact, FCI, energy, from the information contained in nothing more than $\boldsymbol{T_3}$ and $\boldsymbol{T_4}$, as long as these cluster components are extracted from the exact CC, or FCI, wave function.

Of course, this is a great exercise in futility, since there is no point in performing extra calculations once FCI energies and wave functions are known. Moreover, if we disregard $\boldsymbol{T_3}$ and $\boldsymbol{T_4}$ altogether, the aforementioned considerations lead us back to the basic CCSD method, which we are intending to improve. Thus, the goal of the externally corrected methods is to find a non-CC source of $\boldsymbol{T_3}$ and $\boldsymbol{T_4}$, providing clusters of good enough quality to tackle challenging problems dominated by stronger correlation effects. Previous externally corrected CC methods have exploited wave functions such as the projected UHF (PUHF) [205, 206], valence bond type wave functions [229], complete active space SCF [208], and others [118, 209–212, 214].

The CAD-FCIQMC method, introduced in this Ph.D. thesis, estimates the contributions of T_3 and T_4 from the cluster analysis of i-FCIQMC wave functions, which are guaranteed to converge to their corresponding exact, FCI, state at the infinite imaginary-time limit.

To understand how this works, let us begin with the time-dependent i-FCIQMC wave function, given by

$$|\Psi_0^{(MC)}(\tau)\rangle = \left[\mathbb{1} + \sum_{n=1}^{N} C_n^{(MC)}(\tau) \right] |\Phi\rangle = \exp\left[\sum_{n=1}^{N} T_n^{(MC)}(\tau) \right] |\Phi\rangle, \qquad (2.85)$$

where we have imposed the intermediate normalization condition $\langle\Phi|\Psi_0^{(MC)}(\tau)\rangle = 1$, such that the time-dependent $c_k(\tau)$ coefficients, entering the CI operator,

$$C_n^{(MC)}(\tau) = \sum_{k} c_k(\tau) E_k, \qquad (2.86)$$

where, E_k is the elementary particle-hole excitation operator, such that $|\Phi_k\rangle = E_k |\Phi\rangle$, can be equated to the exponential parametrization given by $\exp\left[\sum_{n=1}^{N} T_n^{(MC)}(\tau) \right] |\Phi\rangle$. At any given point during a i-FCIQMC simulation, we can always compute the CI coefficients using the signed walker populations $N_k = \sum_\alpha s_\alpha \delta_{k,k_\alpha}$, defined in Sec. 2.4, by

$$c_k(\tau) = \frac{N_k}{N_0}, \qquad (2.87)$$

where N_0 is the number of walkers residing in the reference determinant, $|\Phi\rangle$.

Once the time-dependent CI coefficients are determined, we exploit the equivalence between the FCI and CC wave functions, which has been noticed even since the seminal work by Čížek [25], and which is given by

$$\begin{aligned}
C_1 &= T_1 \\
C_2 &= T_2 + \tfrac{1}{2} T_1^2 \\
C_3 &= T_3 + T_1 T_2 + \tfrac{1}{6} T_1^3 \\
C_4 &= T_4 + T_1 T_3 + \tfrac{1}{2} T_2^2 + \tfrac{1}{2} T_1^2 T_2 + \tfrac{1}{24} T_1^4, \text{ etc.}
\end{aligned} \qquad (2.88)$$

Clearly, after inverting these relationships, we can determine $T_3^{(\mathrm{MC})}(\tau)$ and $T_4^{(\mathrm{MC})}(\tau)$ clusters components, and we can do so at arbitrary imaginary-time snapshots.

In the next step, we can substitute the higher-than-double components in the projection equations, Eqs. (2.82) and (2.83), to obtain

$$\langle \Phi_i^a | [H_N (\mathbb{1} + T_1 + T_2 + \tfrac{1}{2} T_1^2 + T_3^{(\mathrm{MC})}(\tau) + T_1 T_2 + \tfrac{1}{6} T_1^3)]_C |\Phi\rangle = 0, \tag{2.89}$$

and

$$\begin{aligned}
\langle \Phi_{ij}^{ab} | [H_N (\mathbb{1} + T_1 + T_2 + \tfrac{1}{2} T_1^2 + T_3^{(\mathrm{MC})}(\tau) + T_1 T_2 + \tfrac{1}{6} T_1^3 \\
+ T_4^{(\mathrm{MC})}(\tau) + T_1 T_3^{(\mathrm{MC})}(\tau) + \tfrac{1}{2} T_2^2 + \tfrac{1}{2} T_1^2 T_2 + \tfrac{1}{24} T_1^4)]_C |\Phi\rangle = 0.
\end{aligned} \tag{2.90}$$

which we can solve, iteratively, by letting T_1 and T_2 relax while keeping $T_3^{(\mathrm{MC})}(\tau)$ and $T_4^{(\mathrm{MC})}(\tau)$ frozen. Once the amplitudes defining T_1 and T_2 converge to within suitable numerical threshold, we can compute the resulting CAD-FCIQMC energy by taking the expectation value,

$$E_0^{(\mathrm{CAD\text{-}FCIQMC})} = \langle \Phi | [H_N (T_1 + \tfrac{1}{2} T_1^2 + T_2)]_C |\Phi\rangle . \tag{2.91}$$

51

CHAPTER 3

APPLICATION OF THE SEMI-STOCHASTIC CC(P;Q) AND CAD-FCIQMC APPROACHES TO MOLECULAR MODEL SYSTEMS

3.1 Evaluating the Semi-Stochastic CC(P;Q) Approach by Recovering High-Level Ground-State CC Energetics Characterizing the F_2 and H_2O Molecules, and the Automerization Reaction of Cyclobutadiene

Following the introduction of the semi-stochastic CC(P;Q) methodology, the next priority was to assess the effectiveness of these new considerations. In order to do so, we relied on the well studied and very challenging F_2, H_2O, CH^+, and cyclobutadiene molecules, which have been extensively used to test the performance of QC methodologies [51, 164, 202, 203, 213].

At first glance, the F_2 molecule seems to be quite a simple system. It is a weakly bound ($D_e = 44.924\,mE_h$ at the CCSDT level of theory [50]) homo-nuclear diatomic molecule with only 18 electrons. However, as the single σ bond is stretched from the equilibrium geometry, where the internuclear $F-F$ distance is $R_e = 2.668\,16\,bohr$, toward the dissociation limit, large and mostly dynamical correlation effects begin to appear. The higher-order correlation effects are so large, that indeed, CCSD calculations using the cc-pVDZ basis-set [230] become grossly insufficient, producing a potential energy curve that overbinds the F_2 molecule by almost twice its correct value [50]. However, by including triply excited cluster components, the situation improves considerably. The presence of T_3 reduces the correlation energy at R_e by $9.485\,mE_h$ and to a staggering $49.816\,mE_h$ at $5R_e$. Since the T_3 cluster component contribution to the description of the F_2 potential energy curve is this large, we can use this example to assess the performance of methods that approximate CCSDT. In this section, we will show that the semi-stochastic CC(P;Q) method is able to reproduce CCSDT energies describing the F_2 potential energy curve along bond-breaking coordinates, by quickly

identifying the leading triply excited cluster components and correcting for the missing mostly dynamical-correlation effects, achieving sub-mE_h accuracy at the fraction of the computer cost associated with CCSDT calculations.

Even though the F_2 molecule is a great candidate to test SRCC methods including up to triply excited cluster components, it is not equally effective in assessing the effects of quadruples, since the extra accuracy gains achieved by including T_4 are quite small for this particular system. Thus, in order to explore the robustness of the semi-stochastic CC($P;Q$) methodology when using P spaces spanned by up to quadruply excited determinants, we relied, instead, on the the double bond dissociation of the H_2O molecule. This challenging test case consists in the computation of the ground-state potential energy curve along the dimension corresponding to the simultaneous, C_{2v}-symmetric, stretching of both O−H bonds while maintaining a constant angle of $\angle(\text{H}-\text{O}-\text{H}) = 110.57°$ [231, 232]. At the equilibrium geometry, defined by $R_e = R_{\text{O}-\text{H}} = 1.8434\,\text{bohr}$, the effect of including T_3 on top of CCSD calculations using the cc-pVDZ basis set amounts, already, to $3.251\,\text{m}E_h$. By increasing the CC level to CCSDTQ, the ground-state energy improves by an extra $0.474\,\text{m}E_h$, reaching an error of only $0.019\,\text{m}E_h$ relative to the FCI energy. The effect of triply and quadruply excited cluster components become much larger as stretched geometries are considered. For example, at $R_{\text{O}-\text{H}} = 2.5R_e$, the CCSDT ground-state energy produces a huge error of $-24.752\,\text{m}E_h$ relative to FCI, which is drastically reduced in CCSDTQ calculations, to only $-2.361\,\text{m}E_h$. The huge effect of including T_4 cluster components, which shift the CCSDT correlation energy by about $22\,\text{m}E_h$, makes the water dissociation example a great candidate for testing the efficacy of the semi-stochastic CC($P;Q$) approach in capturing the correlation effects in quadruply excited cluster components.

A more sizable example is the one given by the automerization reaction of cyclobutadiene. This is a highly reactive, 28 electron, anti-aromatic system, containing π-orbitals, and which constitutes a very difficult challenge to QC methods. The reaction consists
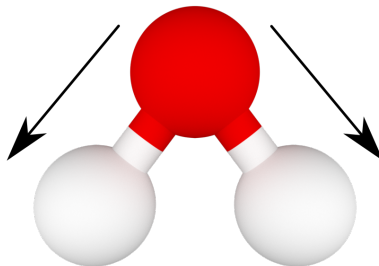
Figure 3.1: Illustration of the water molecule using a ball and stick model. In this work, we employ the semi-stochastic CC(P;Q) approach to calculate ground-state energies entering the potential energy curve generated by the simultaneous bond stretching of both $O-H$ bonds, indicated by the arrows.

in the rearrangement of the π-electrons from a stable configuration in which two opposite sides of a rectangle are held by double bounds, to the other possible closed-shell configuration, where the double bonds are between the remaining sides (see Figure 3.2). The difficulty in describing this process lies in the correct balancing of the electronic correlation in both non-degenerate reactant (or product) state, and the more complicated transition state defined by a strongly quasi-degenerate biradical character. This requires electronic structure methods capable of recovering large dynamical, as well as nondynamical (particularly in the transition state), correlation effects with considerable accuracy. Experimental results on the temperature and concentration dependence of vicinally dideuterated cyclobutadiene, set the limits of the activation energy corresponding to the automerization of cyclobutadiene between 1.6 and 10 kcal/mol [233]. This wide range of possible activation energies provides a great opportunity to employ QC theory in attaining a more robust understanding of this system, which improves upon experimental estimates. Various QC methods have been employed in order to calculate the activation energy of this system, including both SR and MR methods, of CI and CC theories. Unfortunately, SRCC methods with up to doubly excited cluster components, such as CCSD, CCSD(T), and CR-CC(2,3), overestimate the activation energy barriers, producing values that range from 15.8 kcal/mol to 20.9 kcal/mol [164]. These large errors emphasize the difficulty in balancing the nondynamical and dynamical correlation effects, required for
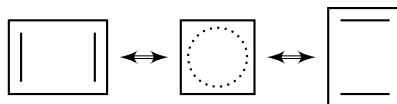
Figure 3.2: Automerization reaction of cyclobutadiene. The structures shown to the left and right correspond to the rectangular geometry that defines both the reactant and product states. The middle structure represents the transition state characterized by the square geometry.

the description of this reaction. When including higher-than-two-body components in the cluster operator, SRCC methods, such as CCSDt, CC(t;3), and the parent CCSDT, estimate the barrier within the experimental bounds, between 7 and 8 kcal/mol [164]. These values are also closer to the ones provided by the much more sophisticated and computationally expensive MR CC methods, such as the MkCCSD(T) [234], CAS-BCCC4 [235], RMRCCSD(T) [236], MRAQCC [237], and SS-EOMCCSD[+2] [238], which locate the activation energy in the 7–9 kcal/mol range. Nevertheless, the mostly dynamical correlation effects missing in CCSDt, still produce large $\sim 20\,mE_h$ errors in the total energies of both reactant and transition state geometries, which corroborates the observation that the automerization of cyclobutadiene requires robust SR methods that can describe large nondynamical and dynamical correlation effects. In summary, since the T_3 effects, defined as the difference between CCSDT and CCSD energies using the cc-pVDZ basis set, are of $26.827\,mE_h$ and $47.979\,mE_h$ for the reactant and transition state geometries, respectively, and because our goal is to approximate CCSDT quality energetics via the semi-stochastic CC(P;Q) approach, the challenging automerization reaction of cyclobutadiene serves as another great example to test our new method's performance.

3.1.1 Computational Details

In the study of all the aforementioned examples, all one- and two-body integrals were obtained from molecular orbitals computed using the RHF SCF methods provided by the popular GAMESS (US) software package [239–241]. Subsequently, the i-FCIQMC simulations, as well as all CC(P) and CC(P;Q) calculations, were performed using these

molecular integrals. The one- and two-body integrals for the F_2 molecule were computed at the geometries $R_e = R_{F-F} = 2.668\,16\,\text{bohr}$, $1.5R_e$, $2R_e$, and $5R_e$ using the cc-pVDZ basis set and the $2R_e$ geometry using the cc-pVTZ and aug-cc-pVTZ basis sets. In the case of the automerization reaction of cyclobutadiene, the integrals corresponding to the rectangular reactant/product geometry, where $R_{C=C} = 1.367\,\text{Å}$, $R_{C-C} = 1.573\,\text{Å}$, $R_{C-H} = 1.093\,\text{Å}$, and $\angle(H-C-C) = 134.9°$; and the square transition state geometry, where $R_{C=C} = 1.461\,\text{Å}$, $R_{C-C} = 1.461\,\text{Å}$, $R_{C-H} = 1.092\,\text{Å}$, and $\angle(H-C-C) = 135.0°$, were both computed using the cc-pVDZ basis set. Both geometries were obtained from Ref. [237], and were optimized at the MR-AQCC/SA-4-CASSCF/cc-pVDZ level of theory. The integrals for the simultaneous bond stretching of the water molecule were computed using the cc-pVDZ basis set for the geometries $R_{O-H} = R_e = 1.8434\,\text{bohr}$, $1.5R_e$, $2R_e$, $2.5R_e$, and $3R_e$ while maintaining $\angle(H-O-H) = 110.57°$, as described in Ref. [242].

The i-CISDT-MC, i-CISDTQ-MC, and i-FCIQMC simulations used in this work were performed using the HANDE stochastic QC package [243, 244]. The i-CISDT-MC and i-CISDTQ-MC runs were achieved by limiting spawning events of the i-FCIQMC algorithm to remain on up to triply or quadruply excited determinants, respectively. The initiator threshold parameter was set to the default, $n_a = 3$, throughout this work. All i-CIQMC calculations pertaining the F_2, H_2O, and cyclobutadiene molecules, were initiated by placing a 100 walkers on the reference, HF, determinant and setting a time step of $\Delta\tau = 0.0001\,\text{a.u.}$ Since the goal of all i-CIQMC simulations is to produce lists of determinants to be used in the definition of the P spaces entering CC($P;Q$) calculations, and, thus, the numerical convergence and statistical analysis of the i-CIQMC simulations was not required, all runs were left in the exponential growth mode, by leaving the shift energy at its initial value corresponding to the HF energy.

The CC codes utilized in all stochastic CC($P;Q$) calculations are homegrown codes, implementing spin-integrated CC and EOMCC methods with up to quadruply and

triply excited components, respectively. The ground-state CC code consists in an itera-tive Jacobi solver assisted by the DIIS procedure [245–247], that is capable of solving the spin-integrated systems of equations characterizing the CCSD, CCSDt, CCSDT, CCSDtq, CCSDTq, and CCSDTQ methods. The CC(P) calculations were carried out by modify-ing the CCSDT subroutines, forcing them to only calculate those triples, or triples and quadruples, contained in the P space. The left-CC equations given by the eigenvalue problem $\langle\Phi|L_0\bar{H}^{(P)}|\Phi_k\rangle = l_k E_0^{(P)}$, and required for the computation of the corrections $\delta(P;Q)$, were solved as an homogeneous linear system of equations since the eigenvalue problem can be rewritten as $\langle\Phi|L_0(\bar{H}^{(P)} - E_0^{(P)}\mathbb{1}|\Phi_k\rangle = 0$ and because the energy $E_0^{(P)}$ is known from the previous ground-state CC calculations. As in the CC case, our left CC codes implement the Jacobi algorithm assisted by DIIS [245–247] for the solution of this system of equations, and were modified according to the aforementioned descrip-tion to compute only those triples belonging to the P space. Also, in all calculations corresponding to the F_2 and cyclobutadiene molecules, the lowest energy core molecular orbitals was kept frozen. In the case of the H_2O, all electrons were correlated.

3.1.2 Numerical Results

The results of the evaluation of the semi-stochastic CC($P;Q$) methods, when applied to the model systems described above, are quite spectacular. As we will see in section, they are able recover their parent's CCSDT and CCSDTQ energies at the very early stages of the QMC propagations.

The results of the semi-stochastic CC($P;Q$) calculations of the bond stretching of F_2/cc-pVDZ molecule from $R_e = 2.668\,16\,\text{bohr}$ to $5R_e$ are shown in Figures 3.3, 3.4, and 3.5 and Table 3.1. The figures show the projective i-CISDT-MC, i-CISDTQ-MC, and i-FCIQMC energies as a function of the time-step iterations with a noisy green curve. On top of that curve, lies a red curve with red circles, which depicts the CC(P) energies computed by solving iterative CC equations in the P space defined as the space spanned

by all singly and doubly excited determinants, and only those triply excited determinants recovered at every $10\,000$ iterations of the corresponding QMC simulation. The black curve with empty squares shows the energies of the CC(P;Q) calculations, which consist in the CC(P) energy, as described by the red curve, corrected by the missing correlation effects contained in the Q space, and described by determinants not captured by the QMC simulations. In all figures the energies are plotted as energy differences relative to CCSDT, in mE_h, since the goal in this example is to recover CCSDT ground-state energies without incurring in the large computational cost of $n_o^3 n_u^5$ CPU time steps. At the 0 iterations point, the QMC projective energy is the uncorrelated RHF energy, since all available walkers are on the reference determinant. On the other hand, at this point the CC(P) and CC(P;Q) energies are equivalent to the ones computed using CCSD and CR-CC(2,3), due to the fact that the P space contains no triply excited components, and, thus, is spanned by only singly and doubly excited determinants.

As the QMC simulations progress, the resulting CC(P) and CC(P;Q) energies rapidly converge toward the CCSDT parent energy. By $60\,000$ iterations of the corresponding i-FCIQMC, i-CISDTQ-MC, or i-CISDT-MC simulations, our CC(P) energies computed at every geometry and basis set are already well within $1\,\text{m}E_\text{h}$ of the CCSDT target, with most of them being closer, at around $500\,\mu E_\text{h}$. This is achieved with only between 40 and $60\,\%$ of the triples. In contrast, at the same stage in the simulation, the projective i-CIQMC energies are still very noisy, fluctuating, in the best cases, within a $10\,\text{m}E_\text{h}$ range. Moreover, the i-FCIQMC- and i-CISDTQ-MC-based CC(P) energies, computed at all geometries, are able to reach sub-mE_h accuracy, earlier, at $50\,000$ iterations, with the best example being the $5R_e$ stretched geometry, where this accuracy is met at only $30\,000$ iterations. This means that with only 15–40 % of triples, one is able to converge CC(P) energies to within $1\,\text{m}E_\text{h}$ of the parent CCSDT method.

Even though our CC(P) results are already very good, supporting the viability of our semi-stochastic approach as an excellent black-box method for recovering high-level CC

energetics at a fraction of the computational cost, the situation improves dramatically when the CC(P;Q) corrections are incorporated. Indeed, our semi-stochastic CC(P;Q) calculations at all geometries and basis-sets yield results with sub-mE_h accuracy relative to CCSDT within 10 000 iterations of the corresponding QMC simulation, with the exception of calculations of F_2/aug-cc-pVTZ which require a little longer simulation time of 20 000 iterations to reach this point. The accumulated amount of triples to reach these values is less than 5 % for all cases (including F_2/aug-cc-pVTZ at $R = 2R_e$), which means that the application of the CC(P;Q) correction improves the efficiency of our CC(P) calculations by about four-fold.

The great performance of the semi-stochastic CC(P;Q) approach relies on the fact that it is quite capable of balancing nondynamical and dynamical correlation effects. At the equilibrium geometry, the description of the F_2 ground-state is dominated by dynamical correlation effects, which, in total, and at CCSDT level of theory, account for 9.485 mE_h relative to the CCSD ground-state energy. The dynamical nature of the correlation effects at this geometry, can be appreciated by the application of the MM-based correction in the CR-CC(2,3) method, which lowers the CCSD error to only -0.240 mE_h. Since T_1 and T_2 cluster amplitudes in the CCSD and CR-CC(2,3) approaches are not able to relax in the presence of higher-order, T_3, amplitudes, they fail to respond to the strong couplings resulting from quasi-degeneracies. Despite this, the CR-CC(2,3) is able to account for much of the dynamical correlation effects in T_3. Nevertheless, as the F_2 bond is stretched, the nondynamical correlation effects become considerably stronger, forcing larger errors on CR-CC(2,3) due to the lack of coupling between T_1 and T_2 and on the other hand T_3, which prevents the relaxation of T_1 and T_2 during the iterative solution of CCSD equations in CR-CC(2,3). The effect of the balance between nondynamical and dynamical correlations can also be observed in the convergence of CC(P) energies. Even though, as discussed above, CC(P) energies fall below the sub-mE_h mark when around 30–40 % of the triples are captured, as the geometries are stretched, the

CC(P) calculations rapidly account for increasingly more correlation energy, during the initial stages of the simulation. For example, after 10 000 iterations and at the equilibrium geometry, CC(P) energies have improved upon CCSD energies only by about 4 mE_h by capturing the correlation effects in 3 % triples. However, at $R = 5R_e$, under the same circumstances, CC(P) energies already account for almost 40 mE_h, which are completely missing from CCSD calculations. This can be thought as if in statical correlation dominated situations, the correlation effects are mostly described by a set of few higher-order cluster components, whereas dynamical correlation effects are distributed among a large set of small-valued cluster amplitudes. At 50 000 iterations, the CC(P;Q) energies are well within the sub-100 μE_h range, being practically exact when compared against CCSDT energies.

Another important observation from our results is the fact that truncated i-CIQMC simulations are as good as the i-FCIQMC simulations in providing the leading triple excitations entering the P spaces characterizing CC(P) and CC(P;Q) calculations. The fraction of triples captured in each of the i-FCIQMC, i-CISDTQ-MC, and i-CISDT-MC remains similar among them as the QMC simulations progress. This happens despite the fact that the quality of the energies resulting from the truncated CI variants are far worse than those calculated using FCIQMC. In fact, all truncated CI methods are not size extensive and are always less accurate than the corresponding level of CC theory. This can be observed in Figures. 3.4 and 3.5, where the green curve representing the projective energies of i-CISDTQ-MC and i-CISDT-MC simulations, respectively, converge toward the wrong value, especially when stretched geometries of F$_2$ are considered. The fact that using i-CISDT-MC is almost as good as i-FCIQMC in our semi-stochastic CC(P;Q) approach, is of remarkable consequences, since we are not interested in the energy values of the QMC approaches, and the computational savings gain associated with limiting the allowed excitations in a QMC simulation is huge.

As we mentioned earlier, we can also substitute the CIQMC simulations with their

CCMC counterparts in order to extract the leading excited components entering our CC(P) calculations. The results of employing i-CCSDT-MC simulations to identify the triply excited components entering the cluster operator $T^{(P)} = T_1 + T_2 + T_3^{(MC)}$ in CC(P) calculations are shown in Table 3.3. As one can readily see, the rapid convergence of the semi-stochastic CC(P) and their corresponding CC($P;Q$) corrections does not change much when compared to the results presented above using i-CIQMC wave functions, and, thus, it suggests that the semi-stochastic CC($P;Q$) methodology is flexible enough to perform well with various different sources of stochastically determined many-electron wave functions.

In contrast to F_2 which contains only 18 electrons, cyclobutadiene is a 28 electron system, with π-electron system, that requires robust QC methods for attaining quantitative descriptions. Regardless, the semi-stochastic CC($P;Q$) performed very well in the description of the automerization reaction of cyclobutadiene, as was the case in the previous example. At 60 000 iterations and with approximately 50 % of triples captured during any of the i-FCIQMC, i-CISDTQ-MC, or i-CISDT-MC simulations, the CC(P) activation energies, defined as the difference between the ground-state energies of the reactant and transition states given by the rectangle and square geometries, and which CCSDT calculations estimate at 7.637 kcal/mol, are already well within the so-called "chemical accuracy" of 1 kcal/mol. If we include the corrections due to the missing triples not captured during the QMC simulations, this accuracy is reached much earlier, at 40 000 iterations and with only about 15 % of triples (this time with the exception of i-CISDT-MC-based CC($P;Q$) energies which remain almost 2 kcal/mol above the CCSDT target). The performance of the semi-stochastic CC(P) method in describing the ground-state energies of the rectangle and square geometries is also quite remarkable. Although 74–85 % triples are required to achieve sub-mE_h accuracy, one has to consider that the effect of T_3 in both geometries is massive, with substantial dynamical correlations, which require a large P space to describe properly. In the case of the reactant state, the mostly dynamical

Table 3.1: Convergence of the CC(P), CC(P;Q)$_{\text{MP}}$, and CC(P;Q)$_{\text{EN}}$ energies toward CCSDT, where the P spaces consisted of all singles and doubles and subsets of triples identified during the *i*-FCIQMC, *i*-CISDTQ-MC, or *i*-CISDT-MC propagations with $\Delta\tau = 0.0001$ a.u. and where the corresponding Q spaces consisted of the triples not captured by the corresponding QMC simulations, for the F_2/cc-pVDZ molecule in which the F–F distance R was set at R_e, $1.5R_e$, $2R_e$, and $5R_e$, with $R_e = 2.66816$ bohr representing the equilibrium geometry. The *i*-FCIQMC, *i*-CISDTQ-MC, and *i*-CISDT-MC calculations preceding the CC(P) and CC(P;Q) steps were initiated by placing 100 walkers on the RHF determinant and the n_a parameter of the initiator algorithm was set at 3. In all post-RHF calculations, the lowest two core orbitals were kept frozen and the Cartesian components of d orbitals were employed throughout.

R/R_e	MC iterations	% of triples FCI[2]	CIQ[3]	CIT[4]	CC(P)[1] FCI[2]	CIQ[3]	CIT[4]	CC(P;Q)$_{\text{MP}}$[1] FCI[2]	CIQ[3]	CIT[4]	CC(P;Q)$_{\text{EN}}$[1] FCI[2]	CIQ[3]	CIT[4]
1.0	0		0			9.485[5]			1.398[6]			−0.240[7]	
	10000	3	3	4	5.692	5.692	5.229	0.760	0.760	0.688	−0.151	−0.151	−0.152
	20000	9	8	8	3.548	3.804	3.962	0.444	0.473	0.472	−0.107	−0.093	−0.140
	30000	15	16	14	2.290	2.498	2.769	0.284	0.301	0.334	−0.059	−0.046	−0.067
	40000	25	26	22	1.791	1.523	1.765	0.212	0.184	0.210	−0.037	−0.030	−0.034
	50000	37	38	34	0.933	0.940	1.151	0.113	0.115	0.137	−0.014	−0.013	−0.021
	60000	51	52	46	0.536	0.498	0.698	0.064	0.058	0.083	−0.008	−0.008	−0.010
	70000	63	64	58	0.383	0.308	0.410	0.044	0.036	0.047	−0.006	−0.004	−0.007
	80000	73	74	68	0.177	0.164	0.224	0.020	0.018	0.025	−0.003	−0.002	−0.003
	100000	89	89	85	0.044	0.050	0.073	0.005	0.006	0.008	0.000	−0.001	−0.001
	120000	97	97	94	0.013	0.010	0.024	0.001	0.001	0.003	0.000	0.000	0.000
	∞		100			−199.102796[8]			—			—	
1.5	0		0			32.424[5]			5.984[6]			1.735[7]	
	10000	3	3	3	14.312	14.220	15.874	2.198	1.980	2.115	0.351	0.321	0.193
	20000	9	8	7	5.589	3.572	5.564	0.629	0.428	0.657	−0.003	−0.000	0.052
	30000	16	18	14	2.728	2.391	2.206	0.323	0.285	0.262	−0.002	0.020	0.021
	40000	27	30	24	1.065	0.706	1.387	0.142	0.084	0.171	0.020	0.009	0.015
	50000	42	45	35	0.482	0.459	0.687	0.062	0.055	0.087	0.009	0.006	0.008
	60000	57	60	49	0.273	0.219	0.336	0.029	0.027	0.041	0.001	0.000	0.005
	70000	70	72	61	0.128	0.106	0.231	0.013	0.011	0.028	0.000	0.000	0.001
	80000	81	82	72	0.064	0.048	0.102	0.006	0.004	0.010	0.000	−0.001	−0.001
	100000	93	94	88	0.012	0.009	0.026	0.001	0.001	0.003	0.000	0.000	0.000
	120000	99	100	96	0.001	0.002	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	∞		100			−199.065882[8]			—			—	
2.0	0		0			45.638[5]			6.357[6]			1.862[7]	
	10000	4	4	3	12.199	17.779	12.687	0.998	1.886	1.181	−0.063	0.280	−0.008
	20000	10	9	9	4.127	2.529	3.672	0.328	0.245	0.310	−0.014	0.009	−0.025
	30000	21	19	17	0.802	1.172	1.393	0.081	0.115	0.128	0.008	0.011	0.004
	40000	35	32	28	0.456	0.499	0.627	0.040	0.047	0.058	−0.001	0.000	0.000
	50000	51	48	41	0.216	0.215	0.305	0.018	0.019	0.027	−0.001	0.000	−0.001
	60000	66	64	56	0.083	0.112	0.160	0.007	0.010	0.014	−0.001	−0.001	−0.001
	70000	79	75	68	0.037	0.048	0.074	0.003	0.004	0.006	0.000	−0.001	−0.001
	80000	87	85	78	0.013	0.019	0.034	0.001	0.002	0.003	0.000	0.000	0.000
	100000	97	95	91	0.001	0.002	0.007	0.000	0.000	0.001	0.000	0.000	0.000
	120000	100	100	98	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	∞		100			−199.058201[8]			—			—	
5.0	0		0			49.816[5]			3.895[6]			1.613[7]	
	10000	3	3	3	10.887	13.326	9.776	0.455	0.672	0.642	−0.005	0.059	0.202
	20000	8	8	8	1.968	2.535	1.315	0.152	0.165	0.102	0.040	0.026	0.012
	30000	17	15	15	0.529	0.752	1.042	0.041	0.056	0.081	0.001	0.006	0.015
	40000	27	26	26	0.295	0.351	0.346	0.022	0.024	0.025	0.001	−0.001	−0.001
	50000	38	37	36	0.116	0.147	0.166	0.008	0.011	0.011	−0.001	0.000	−0.001
	60000	47	46	44	0.047	0.059	0.070	0.003	0.004	0.005	−0.001	0.000	−0.001
	70000	54	52	50	0.016	0.020	0.030	0.001	0.001	0.002	0.000	0.000	0.000
	80000	60	59	55	0.006	0.006	0.014	0.000	0.000	0.001	0.000	0.000	0.000
	100000	74	73	66	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	120000	89	87	78	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	∞		100			−199.058586[8]			—			—	

[1] Unless otherwise stated, all energies are reported as errors relative to CCSDT in millihartree.

[2] FCI stands for *i*-FCIQMC.

[3] CIQ stands for *i*-CISDTQ-MC.

[4] CIT stands for *i*-CISDT-MC.

[5] Equivalent to CCSD.

[6] Equivalent to the CCSD energy corrected for the effects of T_3 clusters using the CCSD(2)$_T$ approach of Ref. [248], which is equivalent to the approximate form of the completely renormalized CR-CC(2,3) approach of Refs. [144, 145], abbreviated sometimes as CR-CC(2,3)$_A$.

[7] Equivalent to the CCSD energy corrected for the effects of T_3 clusters using the most complete variant of the completely renormalized CR-CC(2,3) approach of Refs. [144, 145], abbreviated sometimes as CR-CC(2,3)$_D$.

[8] Total CCSDT energy in hartree.

Figure 3.3: Convergence of the CC(*P*) (red filled circles and dashed lines) and CC(*P*;*Q*)$_{EN}$ (black open squares and solid lines) energies toward CCSDT for the F$_2$/cc-pVDZ molecule in which the F–F distance *R* was set at (a) R_e, (b) 1.5R_e, (c) 2R_e, and (d) 5R_e, where $R_e = 2.668\,16$ bohr is the equilibrium geometry. The *P* spaces consisted of all singles and doubles and subsets of triples identified during the *i*-FCIQMC propagations with $\Delta\tau = 0.0001$ a.u. (depicted by the green lines representing the corresponding projected energies). The *Q* spaces consisted of the triples not captures by *i*-FCIQMC. All energies are errors relative to CCSDT in mE_h and the insets show the percentages of triples captures during the *i*-FCIQMC propagations.

Table 3.2: Convergence of the CC(P), CC($P;Q$)$_{MP}$, and CC($P;Q$)$_{EN}$ energies toward CCSDT, where the P spaces consisted of all singles and doubles and subsets of triples identified during the i-FCIQMC, i-CISDTQ-MC, or i-CISDT-MC propagations with $\Delta\tau = 0.0001$ a.u. and where the corresponding Q spaces consisted of the triples not captured by the corresponding QMC simulations, for the F$_2$ molecule in which the F–F distance R was set at twice the equilibrium bond length, using the cc-pVTZ and aug-cc-pVTZ basis sets, abbreviated as VTZ and AVTZ, respectively. The i-FCIQMC, i-CISDTQ-MC, and i-CISDT-MC calculations preceding the CC(P) and CC($P;Q$) steps were initiated by placing 100 walkers on the RHF determinant and the n_a parameter of the initiator algorithm was set at 3. In all post-RHF calculations, the lowest two core orbitals were kept frozen and the spherical components of d and f orbitals were employed throughout.

Basis set	MC iterations	% of triples			CC(P)[1]			CC($P;Q$)$_{MP}$[1]			CC($P;Q$)$_{EN}$[1]		
		FCI[2]	CIQ[3]	CIT[4]	FCI[2]	CIQ[3]	CIT[4]	FCI[2]	CIQ[3]	CIT[4]	FCI[2]	CIQ[3]	CIT[4]
VTZ	0		0			62.819[5]			9.211[6]			4.254[7]	
	10000	1	1	1	29.714	31.973	31.571	2.738	3.104	2.636	0.728	0.896	0.539
	20000	2	2	2	11.179	14.687	20.194	0.824	1.097	1.487	0.071	0.151	0.217
	30000	6	6	4	5.787	6.031	9.294	0.400	0.425	0.617	0.028	0.030	0.025
	40000	14	14	10	2.406	2.574	4.203	0.160	0.171	0.284	0.002	0.001	0.014
	50000	27	26	19	1.193	1.237	2.177	0.076	0.078	0.138	−0.003	−0.002	−0.002
	60000	42	42	30	0.490	0.489	1.144	0.029	0.029	0.071	−0.002	−0.002	−0.005
	70000	59	57	44	0.178	0.171	0.576	0.011	0.010	0.037	−0.001	−0.001	−0.002
	80000	72	71	56	0.045	0.054	0.309	0.003	0.003	0.020	0.000	0.000	−0.001
	100000	90	89	78	0.002	0.003	0.130	0.000	0.000	0.009	0.000	0.000	0.000
	∞		100			−199.238344[8]			—			—	
AVTZ	0		0			65.036[5]			9.808[6]			5.595[7]	
	10000	0	0	0	36.316	38.874	42.801	3.641	4.144	4.851	1.594	1.786	2.304
	20000	1	1	1	17.190	20.799	26.557	1.276	1.656	2.288	0.382	0.512	0.791
	30000	4	4	3	8.065	9.272	13.279	0.549	0.623	0.928	0.138	0.138	0.246
	40000	10	10	7	4.408	4.677	7.477	0.291	0.307	0.499	0.057	0.062	0.106
	50000	23	22	15	2.208	2.425	3.951	0.136	0.150	0.244	0.016	0.019	0.038
	60000	41	39	27	1.021	1.137	2.052	0.058	0.070	0.124	0.002	0.005	0.013
	70000	61	58	61	0.385	0.455	0.385	0.021	0.025	0.059	0.000	0.000	0.001
	80000	78	76	78	0.125	0.154	0.125	0.007	0.008	0.026	0.000	0.000	0.000
	100000	97	96	97	0.007	0.009	0.007	0.000	0.001	0.004	0.000	0.000	0.000
	∞		100			−199.253022[8]			—			—	

[1] Unless otherwise stated, all energies are reported as errors relative to CCSDT in millihartree.

[2] FCI stands for i-FCIQMC.

[3] CIQ stands for i-CISDTQ-MC.

[4] CIT stands for i-CISDT-MC.

[5] Equivalent to CCSD.

[6] Equivalent to the CCSD energy corrected for the effects of T_3 clusters using the CCSD(2)$_T$ approach of Ref. [248], which is equivalent to the approximate form of the completely renormalized CR-CC(2,3) approach of Refs. [144, 145], abbreviated sometimes as CR-CC(2,3)$_A$.

[7] Equivalent to the CCSD energy corrected for the effects of T_3 clusters using the most complete variant of the completely renormalized CR-CC(2,3) approach of Refs. [144, 145], abbreviated sometimes as CR-CC(2,3)$_D$.

[8] Total CCSDT energy in hartree.

Table 3.3: Convergence of the CC(P), CC($P;Q$)$_{\text{MP}}$, and CC($P;Q$)$_{\text{EN}}$ energies toward CCSDT, where the P spaces consisted of all singles and doubles and subsets of triples identified during the i-FCIQMC or i-CCSDT-MC propagations with $\Delta\tau = 0.0001$ a.u. and where the corresponding Q spaces consisted of the triples not captured by i-FCIQMC or i-CCSDT-MC, for the F_2/cc-pVDZ molecule in which the F–F distance R was set at R_e, $1.5R_e$, $2R_e$, and $5R_e$ where $R_e = 2.66816$ bohr is the equilibrium geometry. The i-FCIQMC and i-CCSDT-MC calculations preceding the CC(P) and CC($P;Q$) steps were initiated by placing 100 walkers (i-FCIQMC) or excips (i-CCSDT-MC) on the RHF determinant and the n_a parameter of the initiator algorithm was set at 3. In all post-RHF calculations, the lowest two orbitals were kept frozen and the Cartesian components of d orbitals were emloyed throughout. All calculations were performed on Dell Precision T-1700 workstation equipped with a quad-core Intel® Core$^{\text{TM}}$ i5-4690 processor at 3.5 GHz using all four cores.

R/R_e	MC iterations	% of triples		CC(P)[1]		CC($P;Q$)$_{\text{MP}}$[1]		CC($P;Q$)$_{\text{EN}}$[1]	
		i-FCIQMC	i-CCSDT-MC	i-FCIQMC	i-CCSDT-MC	i-FCIQMC	i-CCSDT-MC	i-FCIQMC	i-CCSDT-MC
1.0	0		0	9.485[2]		1.398[3]		−0.240[4]	
	10000	3	6	5.692	5.474	0.757	0.732	−0.160	−0.155
	20000	9	12	3.548	3.253	0.463	0.402	−0.084	−0.099
	30000	15	19	2.290	2.610	0.273	0.334	−0.074	−0.043
	40000	25	31	1.791	1.329	0.205	0.161	−0.048	−0.018
	50000	37	38	0.933	0.880	0.109	0.099	−0.020	−0.013
	60000	51	56	0.536	0.505	0.057	0.058	−0.017	−0.005
	70000	63	65	0.383	0.316	0.045	0.037	−0.005	−0.002
	80000	73	76	0.177	0.153	0.020	0.017	−0.002	−0.001
	100000	89	90	0.044	0.042	0.005	0.005	−0.001	0.000
	120000	97	98	0.013	0.013	0.001	0.001	0.000	0.000
	∞		100	−199.102796[5]		—		—	
1.5	0		0	32.424[2]		5.984[3]		1.735[4]	
	10000	3	5	14.312	16.042	2.813	2.994	1.056	1.155
	20000	9	20	5.589	4.006	0.987	0.752	0.397	0.319
	30000	16	34	2.728	1.072	0.436	0.149	0.126	0.037
	40000	27	48	1.065	0.399	0.129	0.035	0.004	−0.006
	50000	42	62	0.482	0.204	0.045	0.019	−0.009	−0.001
	60000	57	75	0.273	0.105	0.028	0.011	0.000	0.001
	70000	70	86	0.128	0.045	0.015	0.005	0.001	0.001
	80000	81	92	0.064	0.020	0.007	0.001	0.001	−0.001
	100000	93	99	0.012	0.004	0.001	0.000	0.000	0.000
	120000	99	100	0.001	0.001	0.000	0.000	0.000	0.000
	∞		100	−199.065882[5]		—		—	
2.0	0		0	45.638[2]		6.357[3]		1.862[4]	
	10000	4	4	12.199	15.810	1.887	2.641	0.915	1.303
	20000	10	15	4.127	2.436	0.596	0.344	0.279	0.150
	30000	21	32	0.802	0.894	0.067	0.099	−0.009	0.023
	40000	35	53	0.456	0.248	0.036	0.014	−0.007	−0.008
	50000	51	64	0.216	0.172	0.010	0.009	−0.011	−0.006
	60000	66	71	0.083	0.095	0.004	0.004	−0.004	−0.005
	70000	79	81	0.037	0.045	0.002	0.003	−0.001	−0.001
	80000	87	89	0.013	0.016	0.001	0.001	−0.001	−0.001
	100000	97	97	0.001	0.003	0.000	0.000	0.000	0.000
	120000	100	100	0.000	0.000	0.000	0.000	0.000	0.000
	∞		100	−199.058201[5]		—		—	
5.0	0		0	49.816[2]		3.895[3]		1.613[4]	
	10000	3	4	10.887	6.169	1.181	0.816	0.752	0.547
	20000	8	14	1.968	1.612	0.256	0.181	0.147	0.096
	30000	17	23	0.529	0.881	0.018	0.080	−0.025	0.035
	40000	27	36	0.295	0.206	0.013	0.013	−0.010	−0.002
	50000	38	46	0.116	0.065	0.005	0.002	−0.004	−0.003
	60000	47	55	0.047	0.030	0.002	0.002	−0.003	−0.001
	70000	54	69	0.016	0.009	0.001	0.000	−0.001	0.000
	80000	60	75	0.006	0.003	0.000	0.000	0.000	0.000
	100000	74	85	0.000	0.000	0.000	0.000	0.000	0.000
	120000	89	94	0.000	0.000	0.000	0.000	0.000	0.000
	∞		100	−199.058586[5]		—		—	

[1] Unless otherwise stated, all energies are reported as errors relative to CCSDT in millihartree.

[2] Equivalent to CCSD.

[3] Equivalent to the CCSD energy corrected for the effects of T_3 clusters using the CCSD(2)$_{\text{T}}$ approach of [248], which is equivalent to the approximate form of the completely renormalized CR-CC(2,3) approach of [144, 145], abbreviated sometimes as CR-CC(2,3)$_{\text{A}}$.

[4] Equivalent to the CCSD energy corrected for the effects of T_3 clusters using the most complete variant of the completely renormalized CR-CC(2,3) approach of [144, 145], abbreviated sometimes as CR-CC(2,3)$_{\text{D}}$.

[5] Total CCSDT energy in hartree.
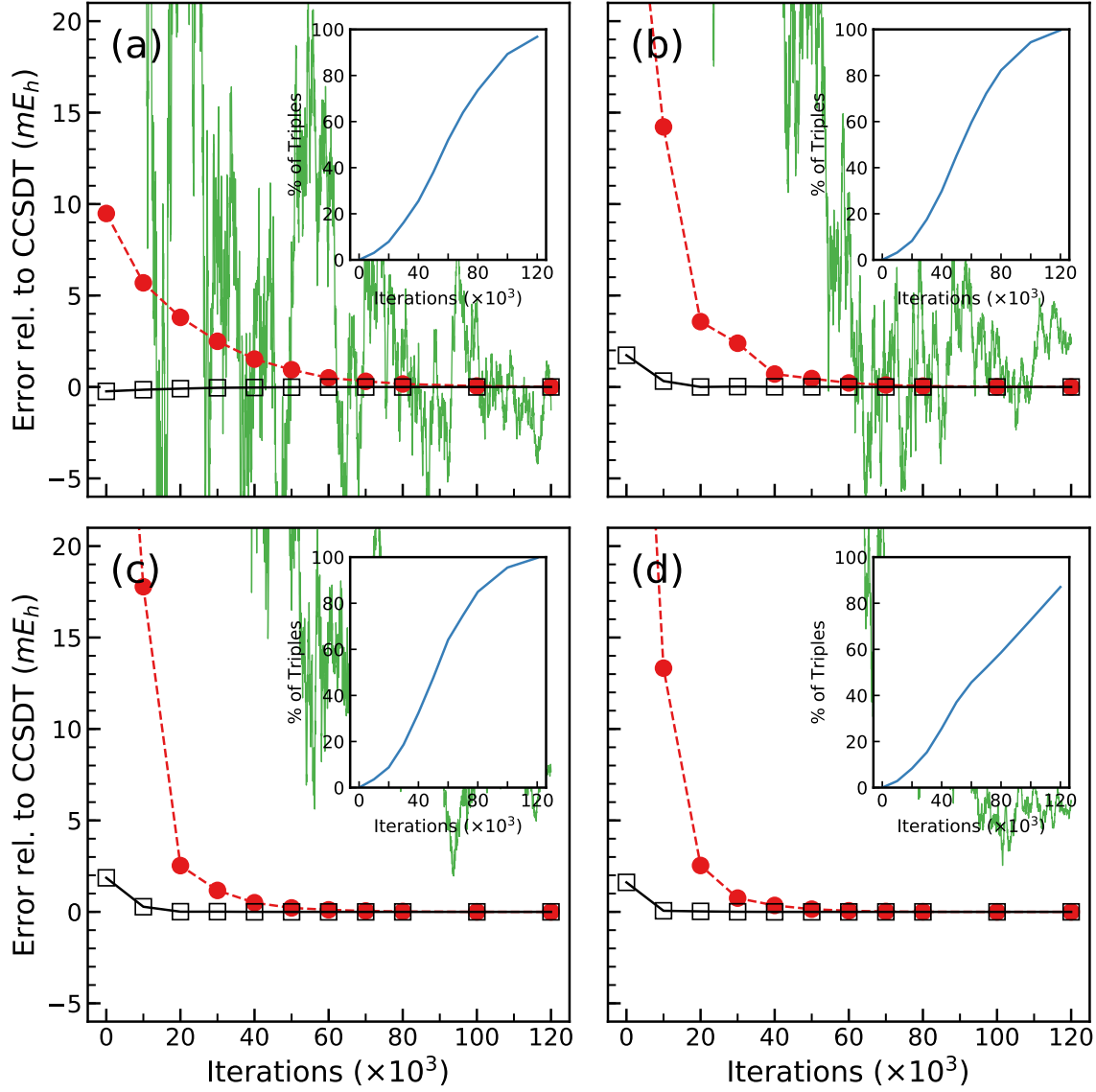
Figure 3.4: Same as Fig. 3.3 except that the subsets of triples included in the CC(*P*) calculations are now identified by the *i*-CISDTQ-MC simulations and the corresponding *Q* spaces consist of the triples not captures by *i*-CISDTQ-MC.
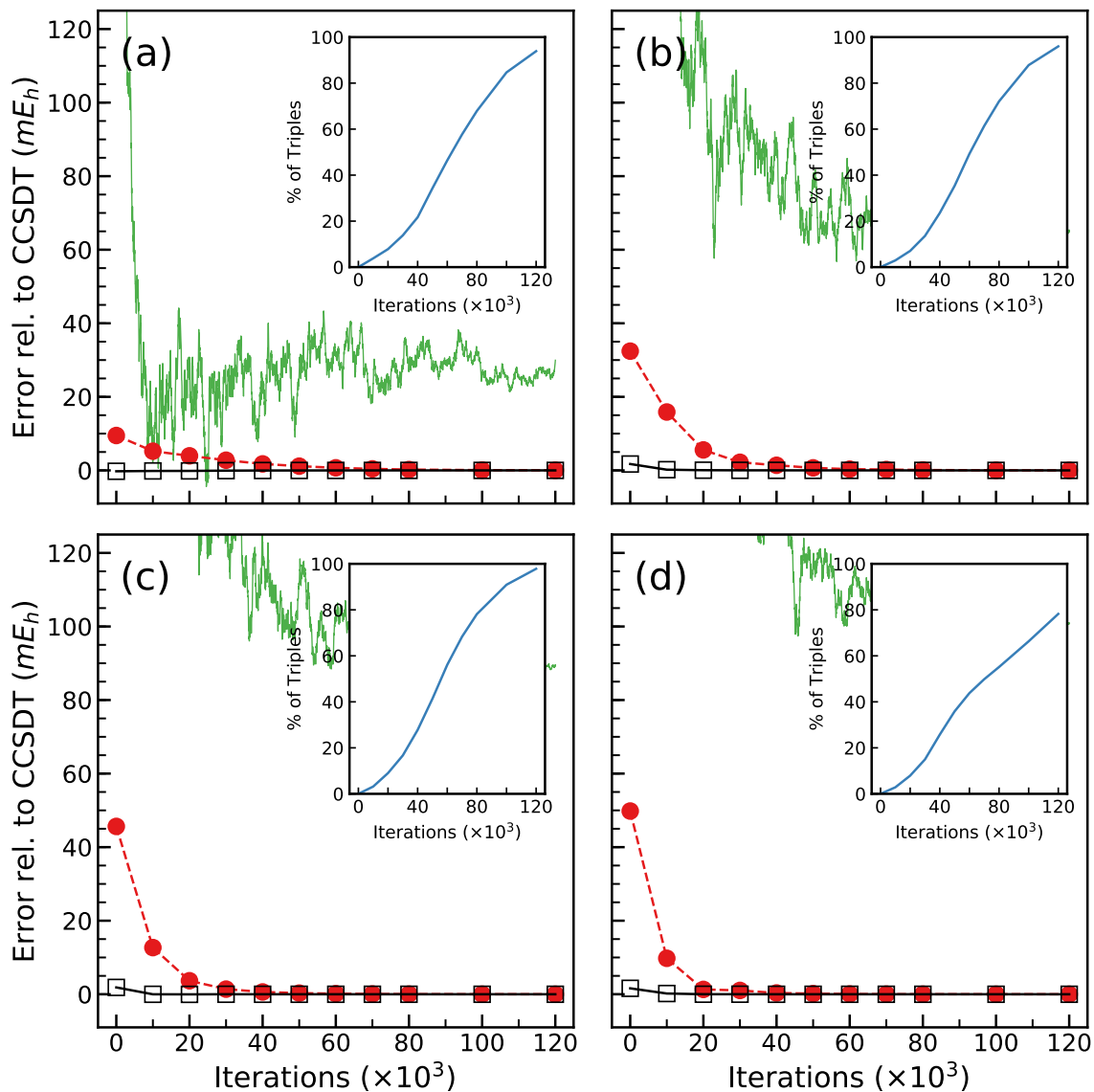
Figure 3.5: Same as Fig. 3.3 except that the subsets of triples included in the CC(P) calculations are now identified by the *i*-CISDT-MC simulations and the corresponding Q spaces consist of the triples not captures by *i*-CISDT-MC.

correlation energy described by triples is 26.827 mE_h while at the reactant state, where a larger multi-reference situation develops, it amounts to 47.979 mE_h. When applying the CC(P;Q) corrections, the energies improve considerably, as expected from our earlier observations from the F$_2$ example. Even before the semi-stochastic CC(P;Q) begins to acquire higher-than-doubly excited components, at 0 iterations when it is equivalent to the CR-CC(2,3), the CC(P;Q) energy reduces the large 26.827 mE_h error of CCSD, at the reactant geometry, to merely 0.848 mE_h. This suggests that most of the correlation energy due to T_3 is of dynamical character at this geometry. As the QMC simulations progress, the semi-stochastic CC(P;Q) energies improve upon the already great performance of CR-CC(2,3) eventually reaching energies below 0.1 mE_h relative to CCSDT, with only 24–31 % of the triples. In the case of the transition state geometry, the inclusion of triples corrections on top of bare CCSD calculations reduce the error by almost 25 mE_h, dropping the correlation energy error from 47.979 mE_h to 14.636 mE_h relative to the CCSDT ground-state energy. The remaining error, mostly due to nondynamical correlations, is corrected by the semi-stochastic CC(P;Q) when enought triples are accumulated in the P space. Within only 40 000 iteration, in which QMC accrued around 15 % of the available triples, the CC(P;Q) energies corresponding to the P spaces build from i-FCIQMC and i-CISDTQ-MC simulations, already reach the sub-mE_h mark. These results show that even in larger and more challenging systems, the semi-stochastic CC(P;Q) methodology is surprisingly robust.

Once the relevance of the semi-stochastic CC(P;Q) approaches was proven by the results presented above, in which we successfully recovered CCSDT-quality energetics using P spaces with only fractions of the total triply excited components, we proceeded to test the performance of the semi-stochastic CC(P;Q) ideas on larger P spaces, spanned with up to quadruply excited Slater determinants. As we discussed before, the simple 10 electron H$_2$O molecule serves as a great example for testing our method, since the contribution of both triply and quadruply excited cluster components to the correlation

Table 3.4: Convergence of the CC(P), CC(P;Q)$_{MP}$, and CC(P;Q)$_{EN}$ energies toward CCSDT, where the P spaces consisted of all singles and doubles and subsets of triples identified during the i-FCIQMC, i-CISDTQ-MC, or i-CISDT-MC propagations with $\Delta\tau = 0.0001$ a.u. and where the corresponding Q spaces consisted of the triples not captured by the corresponding QMC simulations, for the reactant (R) and transition state (TS) structures[1] defining the automerization of cyclobutadiene, as described by the cc-pVDZ basis set, and for the corresponding activation barrier. The i-FCIQMC, i-CISDTQ-MC, and i-CISDT-MC calculations preceding the CC(P) and CC(P;Q) steps were initiated by placing 100 walkers on the RHF determinant and the n_a parameter of the initiator algorithm was set at 3. In all post-RHF calculations, the lowest four core orbitals were kept frozen and the spherical components of d orbitals were employed throughout.

Species	MC iter.	% of triples			CC(P)[2]			CC(P;Q)$_{MP}$[2]			CC(P;Q)$_{EN}$[2]		
		FCI[3]	CIQ[4]	CIT[5]	FCI[3]	CIQ[4]	CIT[5]	FCI[3]	CIQ[4]	CIT[5]	FCI[3]	CIQ[4]	CIT[5]
R	0		0			26.827[6]			4.764[7]			0.848[8]	
	10000	0	0	0	25.758	25.985	25.484	4.437	4.535	4.324	0.696	0.763	0.625
	20000	2	2	1	22.532	22.513	22.462	3.684	3.621	3.612	0.496	0.418	0.433
	30000	6	5	5	17.369	17.857	18.880	2.599	2.676	2.889	0.230	0.228	0.279
	40000	16	15	12	11.845	12.034	13.834	1.635	1.649	2.007	0.092	0.080	0.164
	50000	31	30	24	6.895	7.176	9.202	0.877	0.913	1.235	0.022	0.023	0.057
	60000	52	51	41	3.273	3.524	5.205	0.386	0.417	0.645	0.001	0.000	0.010
	70000	72	70	59	1.321	1.498	2.594	0.146	0.170	0.302	−0.003	−0.002	−0.003
	80000	85	84	75	0.512	0.563	1.181	0.056	0.060	0.131	−0.001	−0.001	−0.002
	∞		100			−154.244157[9]			—			—	
TS	0		0			47.979[6]			20.080[7]			14.636[8]	
	10000	0	0	0	45.875	46.427	45.777	18.899	19.135	18.037	13.680	13.842	12.665
	20000	1	2	1	39.577	37.689	39.655	14.220	12.522	13.774	9.452	7.793	8.863
	30000	5	5	5	30.836	28.405	33.111	9.660	7.404	10.798	5.785	3.648	6.651
	40000	15	13	13	18.976	19.811	23.797	4.046	4.313	6.457	1.556	1.661	3.367
	50000	31	27	26	9.795	9.727	12.495	1.634	1.488	2.238	0.309	0.243	0.602
	60000	52	47	42	3.936	4.136	6.217	0.501	0.525	0.886	0.026	0.025	0.105
	70000	70	67	60	1.491	1.488	2.841	0.173	0.168	0.363	0.003	0.001	0.018
	80000	84	82	74	0.525	0.591	1.260	0.058	0.065	0.148	0.000	0.000	0.001
	∞		100			−154.232002[9]			—			—	
Barrier[7]	0		0/0			13.274[6]			9.611[7]			8.653[8]	
	10000	0/0	0/0	0/0	12.624	12.828	12.734	9.075	9.162	8.605	8.148	8.208	7.555
	20000	2/1	2/2	1/1	10.696	9.523	10.789	6.612	5.586	6.377	5.620	4.628	5.290
	30000	6/5	5/5	5/5	8.450	6.619	8.931	4.431	2.967	4.963	3.487	2.146	3.999
	40000	16/15	15/13	12/13	4.475	4.881	6.252	1.513	1.672	2.793	0.919	0.992	2.011
	50000	31/31	30/27	24/26	1.820	1.601	2.067	0.475	0.361	0.629	0.181	0.138	0.343
	60000	52/52	51/47	41/42	0.416	0.384	0.635	0.073	0.068	0.151	0.016	0.016	0.060
	70000	72/70	70/67	59/60	0.107	−0.006	0.155	0.017	−0.001	0.038	0.003	0.002	0.013
	80000	85/84	84/82	75/74	0.008	0.018	0.050	0.001	0.003	0.011	0.001	0.001	0.002
	∞		100/100			7.637[10]			—			—	

[1] Structures optimized in the MR-AQCC calculations reported in Ref. [237].
[2] Unless otherwise stated, all energies are reported as errors relative to CCSDT, in millihartree for the reactant and transition state and in kcal/mol for the activation barrier.
[3] FCI stands for i-FCIQMC.
[4] CIQ stands for i-CISDTQ-MC.
[5] CIT stands for i-CISDT-MC.
[6] Equivalent to CCSD.
[7] Equivalent to the CCSD energy corrected for the effects of T_3 clusters using the CCSD(2)$_T$ approach of Ref. [248], which is equivalent to the approximate form of the completely renormalized CR-CC(2,3) approach of Refs. [144, 145], abbreviated sometimes as CR-CC(2,3)$_A$.
[8] Equivalent to the CCSD energy corrected for the effects of T_3 clusters using the most complete variant of the completely renormalized CR-CC(2,3) approach of Refs. [144, 145], abbreviated sometimes as CR-CC(2,3)$_D$.
[9] Total CCSDT energy in hartree.
[10] The CCSDT activation barrier in kcal/mol.
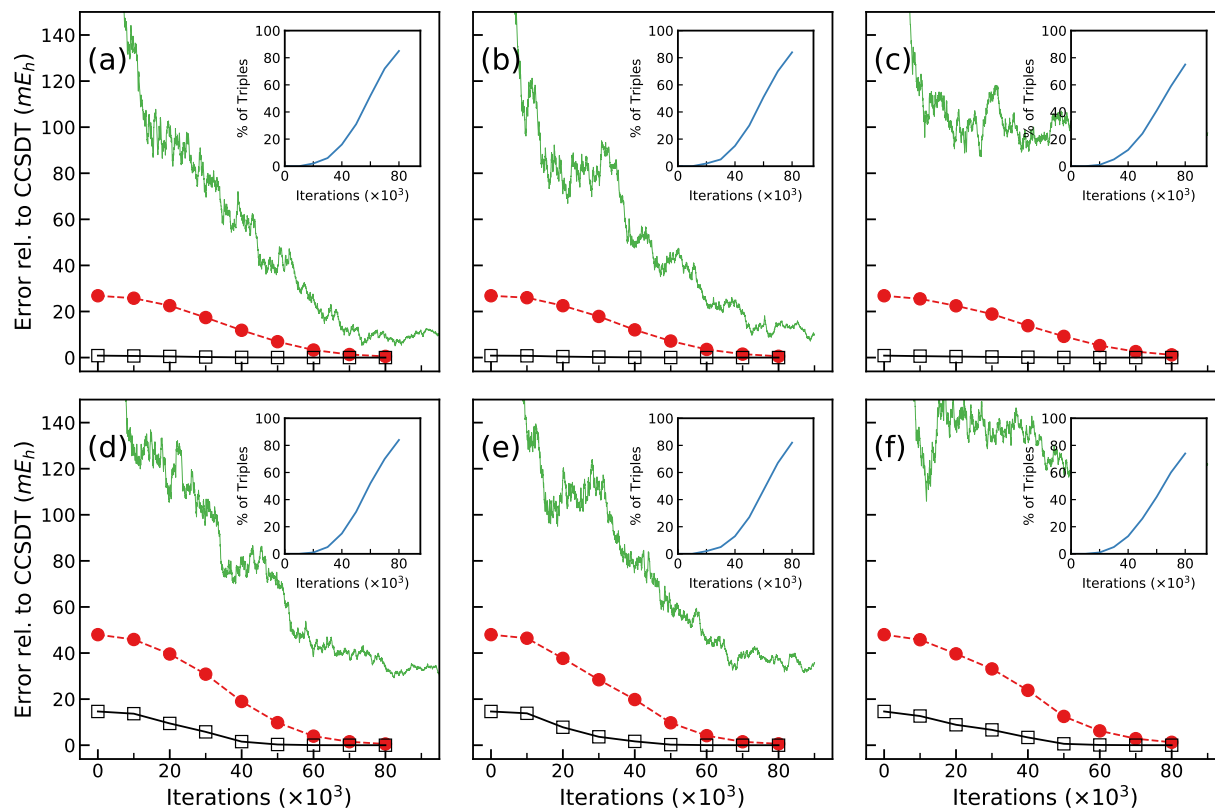
Figure 3.6: Convergence of the CC(P) (red filled circles and dashed lines) and CC(P;Q)$_{\text{EN}}$ (black open squares and solid lines) energies toward CCSDT for the reactant [panels (a)-(c)] and transition state [panels (d)-(f)] structures defining the automerization of cyclobutadiene, as described by the cc-pVDZ basis set. The relevant i-CIQMC runs (all using $\Delta\tau = 0.0001$ a.u.) are depicted by the green lines representing the corresponding projected energies. Panel (a) and (d) correspond to the calculations in which the P spaces employed in the CC(P) steps consisted of all singles and doubles and subsets of triples identified during the i-FCIQMC propagations; the Q spaces needed to define the corresponding $\delta(P;Q)$ corrections consisted of the triples that were not captures by i-FCIQMC. Panels (b) and (e) correspond to the calculations in which the P spaces employed in the CC(P) steps consisted of all singles and doubles and subsets of triples identified during the i-CISDTQ-MC propagations; in this case, the Q spaces needed to define the $\delta(P;Q)$ corrections consisted of the triples that were not captures by i-CISDTQ-MC. Panels (c) and (f) correspond to the calculations in which the P spaces employed in the CC(P) steps consisted of all singles and doubles and subsets of triples identified during the i-CISDT-MC propagations; in this case, the Q spaces needed to define the $\delta(P;Q)$ corrections consisted of the triples that were not captures by i-CISDT-MC. All reported energies are errors relative to CCSDT in mE_{h}. The insets show the percentages of triples captures during the relevant i-CIQMC propagations.

70

energy become quite sizeable, particularly as the simultaneous bond stretching geometries drift far from equilibrium.

At the equilibrium geometry, where $\angle(\text{H}-\text{O}-\text{H}) = 110.57°$ and $R_e = R_{\text{O}-\text{H}} = 1.8434$ bohr, the effect due to triples defined as the difference between CCSD and CCSDT energies amounts to $3.251\,\text{m}E_\text{h}$ while the effect of quadruples, estimated as the difference between CCSDT and CCSDTQ energies, reaches $0.474\,\text{m}E_\text{h}$. As the O$-$H bond lengths are stretched, the effect of triples and quadruples begin to dramatically increase. At the most stretched $R_{\text{O}-\text{H}} = 3R_e$ geometry, the effect of triples and quadruples is staggering. The CCSDT ground-state energy, lowers the correlation energy by a massive $50.975\,\text{m}E_\text{h}$ relative to CCSD, which CCSDTQ improves significantly by increasing it $35.393\,\text{m}E_\text{h}$ to $-75.916\,679\,E_\text{h}$ which is still $-4.733\,\text{m}E_\text{h}$ away from the FCI value. The results of the semi-stochastic CC(P;Q) approach for geometries R_e, $1.5R_e$, $2R_e$, $2.5R_e$, and $3R_e$ are shown in Table 3.6. As has been true in the F$_2$ and cyclobutadiene examples, the CC(P)-calculated ground-state energies, this time exploiting $T = T_1 + T_2 + T_3^{(\text{MC})} + T_4^{(\text{MC})}$, rapidly converge toward the parent CCSDTQ energies for all geometries. At the equilibrium geometry, where the correlation effects are mostly dynamical, the i-FCIQMC and i-CISDTQ-MC methods require at least $80\,000$ iterations to recover $27\,\%$ of the triples and $6\,\%$ of quadruples, which is enough to produce CC(P) ground state energies with sub-mE_h errors. As we examine the increasingly stretched geometries, the sub-mE_h mark is achieved earlier, with only $60\,000$ iterations for $1.5R_e$ and $2R_e$ geometries, $50\,000$ iterations for $2.5R_e$, and $40\,000$ iterations for $3R_e$. As was the case with the equilibrium geometry, the required triples to achieve this accuracy amount to below $30\,\%$ while only less than $6\,\%$ quadruples are needed. The CC(P;Q) approach, in which corrections due to triples not captured during i-CIQMC simulations are included, improves the CC(P) energies substantially. For geometries ranging from R_e to $2R_e$, the CR-CC(2,3) energies are already around or below $1\,\text{m}E_\text{h}$. The semi-stochastic CC(P;Q) energies improve upon this already good value by reducing the errors at only $10\,000$ iterations to

71

Table 3.5: Comparison of CCSD, CCSDT, CCSDTQ ground-state energies for the geometries defined by the O−H bond lengths of R_e, $1.5R_e$, $2R_e$, $2.5R_e$, and $3R_e$ of the simultaneous bond stretching of the H_2O molecule.

Method	$R_{O-H} = R_e$	$R_{O-H} = 1.5R_e$	$R_{O-H} = 2R_e$	$R_{O-H} = 2.5R_e$	$R_{O-H} = 3R_e$
CCSD	3.725	9.922	22.002	22.668	15.582
CCSDT	0.474	1.302	−1.435	−22.391	−35.393
CCSDTQ	−76.241841	−76.072227	−75.951635	−75.920352	−75.916679

within $0.5\,mE_h$. At $2.5R_e$ and $3R_e$, the error accrued by the CR-CC(2,3) energies is rather large, being $-20.739\,mE_h$ and $-35.823\,mE_h$, respectively, which are similar to the errors produced by CCSDT at those geometries. Nonetheless, the CC(P;Q) energies improve considerably. Within $20\,000$ iterations, and with only 5–6 % triples and 1 % quadruples, the semi-stochastic CC(P;Q) energies are able to reduce the large errors of CR-CC(2,3) to sub-mE_h accuracies. These are improvements of about $20\,mE_h$ and $35\,mE_h$ for the $2.5R_e$ and $3R_e$ geometries, and shows the power of the semi-stochastic CC(P;Q) methodology.

As it is now evident from the results and discussions of this section, the performance of semi-stochastic CC(P;Q) methodologies are extremely encouraging. We have shown how the semi-stochastic CC(P;Q) approach, where $T^{(P)} = T_1 + T_2 + T_3^{(MC)}$ during the iterative CC(P) stage, and where the Q space is spanned by all triply excited Slater determinants not captured by i-CIQMC, is able to attain sub-mE_h accuracies at very early stages of the corresponding i-CIQMC simulations, despite the challenging correlation effects involved in the stretching of the F_2 bond and automerization of cyclobutadiene that require robust enough methods to properly balance dynamical and statical correlation effects. The same impressive performance is observed when applying the semi-stochastic CC(P;Q) approach to the simultaneous bond stretching of the H_2O molecule. In this example, the iterative semi-stochastic CC(P) calculations are carried out using $T^{(P)} = T_1 + T_2 + T_3^{(MC)} + T_4^{(MC)}$, and then corrected to account for the effects contained in only triply excited determinants. The semi-stochastic CC(P;Q) energies are able to reduce the large errors of CR-CC(2,3) and even CCSDT ground-state energies

Table 3.6: Convergence of the CC(P), CC(P;Q)$_{\mathrm{MP}}$, and CC(P;Q)$_{\mathrm{EN}}$ energies toward CCS-DTQ, where the P spaces consisted of all singles and doubles and subsets of triples and quadruples identified during the i-FCIQMC or i-CISDTQ-MC propagations with $\Delta\tau = 0.0001$ a.u. and where the corresponding Q spaces consisted of the triples not captured by the corresponding QMC simulations, for the equilibrium and four displaced geometries of the H$_2$O molecule, as described by the cc-pVDZ basis set. The i-FCIQMC and i-CISDTQ-MC calculations preceding the CC(P) and CC(P;Q) steps were initiated by placing 100 walkers on the RHF determinant and the n_a parameter of the initiator algorithm was set at 3. All electrons were correlated and the spherical components of d orbitals were employed throughout.

R/R_e[1]	MC iterations	% of triples/quadruples		CC(P)[2]		CC(P;Q)$_{\mathrm{MP}}$[2]		CC(P;Q)$_{\mathrm{EN}}$[2]	
		FCI[3]	CIQ[4]	FCI[3]	CIQ[4]	FCI[3]	CIQ[4]	FCI[3]	CIQ[4]
1.0	0	0/0	0/0	3.725[5]		0.887[6]		0.325[7]	
	10000	2/0	2/0	3.291	3.291	0.718	0.718	0.220	0.220
	20000	4/1	4/1	2.874	2.874	0.633	0.629	0.205	0.185
	30000	6/1	5/1	2.637	2.637	0.544	0.600	0.143	0.184
	40000	11/2	9/2	2.052	2.052	0.441	0.471	0.142	0.129
	50000	13/2	14/3	1.910	1.910	0.390	0.358	0.105	0.095
	60000	17/3	18/4	1.481	1.481	0.304	0.323	0.087	0.106
	70000	22/5	22/5	1.238	1.238	0.245	0.249	0.065	0.076
	80000	27/6	27/6	0.956	0.956	0.207	0.216	0.073	0.082
	100000	36/10	35/10	0.586	0.586	0.127	0.143	0.048	0.065
	∞		100	−76.241841[8]		—		—	
1.5	0	0/0	0/0	9.922[5]		2.704[6]		1.021[7]	
	10000	3/1	3/1	6.612	6.545	1.393	1.501	0.290	0.434
	20000	8/1	7/1	4.068	4.168	0.898	0.799	0.236	0.138
	30000	11/2	11/2	3.000	3.032	0.613	0.698	0.144	0.248
	40000	16/3	17/3	1.878	2.207	0.481	0.503	0.231	0.189
	50000	22/4	22/4	1.465	1.507	0.377	0.366	0.185	0.166
	60000	26/6	27/6	0.993	0.959	0.254	0.270	0.133	0.152
	70000	31/8	33/9	0.786	0.706	0.229	0.206	0.133	0.122
	80000	36/10	38/11	0.552	0.548	0.186	0.156	0.130	0.091
	100000	46/17	48/18	0.259	0.263	0.086	0.086	0.061	0.060
	∞		100	−76.072227[8]		—		—	
2.0	0	0/0	0/0	22.002[5]		3.775[6]		−0.581[7]	
	10000	2/0	2/0	11.766	11.803	1.966	2.189	−0.044	0.200
	20000	7/1	6/1	4.172	4.937	1.129	1.295	0.567	0.626
	30000	10/2	9/1	3.132	3.788	0.708	0.683	0.323	0.160
	40000	14/3	13/2	1.728	1.966	0.603	0.668	0.436	0.483
	50000	19/4	19/4	1.123	1.120	0.421	0.509	0.324	0.437
	60000	25/6	24/6	0.794	0.719	0.305	0.221	0.246	0.156
	70000	30/8	30/8	0.429	0.427	0.129	0.144	0.094	0.110
	80000	36/11	35/11	0.327	0.293	0.106	0.103	0.079	0.082
	100000	47/18	47/18	0.107	0.102	0.036	0.026	0.029	0.021
	∞		100	−75.951635[8]		—		—	
2.5	0	0/0	0/0	22.668[5]		−13.469[6]		−20.739[7]	
	10000	3/0	3/0	18.305	−3.327	−1.136	−18.549	−4.962	−21.357
	20000	6/1	6/1	5.254	7.207	0.010	0.448	−0.821	−0.588
	30000	10/2	9/2	2.278	2.109	0.513	0.988	0.298	0.872
	40000	15/3	13/3	1.021	1.170	0.304	0.542	0.220	0.490
	50000	22/5	17/4	0.459	0.585	0.264	0.287	0.254	0.264
	60000	27/8	23/6	0.340	0.424	0.105	0.222	0.096	0.212
	70000	34/12	29/9	0.133	0.411	0.059	0.020	0.054	−0.033
	80000	42/16	36/13	0.088	0.155	0.014	0.052	0.011	0.045
	100000	55/28	49/22	0.020	0.027	0.013	0.020	0.012	0.020
	∞		100	−75.920352[8]		—		—	
3.0	0	0/0	0/0	15.582[5]		−28.302[6]		−35.823[7]	
	10000	3/1	3/1	10.165	12.515	−2.390	−1.199	−3.945	−2.697
	20000	5/1	5/1	4.282	2.721	−0.084	−0.690	−0.403	−0.875
	30000	9/2	8/2	1.616	3.019	0.544	0.357	0.414	0.007
	40000	13/3	11/3	0.969	0.830	0.267	0.378	0.199	0.334
	50000	18/5	17/5	0.523	0.400	0.251	0.196	0.231	0.184
	60000	24/8	22/7	0.185	0.237	0.097	0.093	0.090	0.087
	70000	30/12	28/10	0.082	0.128	0.039	0.076	0.036	0.075
	80000	36/16	34/14	0.030	0.050	0.022	0.030	0.021	0.029
	100000	51/28	48/24	0.005	0.012	0.005	0.008	0.005	0.008
	∞		100	−75.916679[8]		—		—	

[1] The equilibrium geometry, $R_{\mathrm{O-H}} = R_e$, and the geometries that represent a simultaneous stretching of both O–H bonds by factors of 1.5, 2.0, 2.5, and 3.0 without changing the $\angle$(H–O–H) angle, were taken from Ref. [249].
[2] Unless otherwise stated, all energies are reported as errors relative to CCSDTQ in millihartree.
[3] FCI stands for i-FCIQMC.
[4] CIQ stands for i-CISDTQ-MC.
[5] Equivalent to CCSD.
[6] Equivalent to the CCSD energy corrected for the effects of T_3 clusters using the CCSD(2)$_{\mathrm{T}}$ approach of Ref. [248], which is equivalent to the approximate form of the completely renormalized CR-CC(2,3) approach of Refs. [144, 145], abbreviated sometimes as CR-CC(2,3)$_{\mathrm{A}}$.
[7] Equivalent to the CCSD energy corrected for the effects of T_3 clusters using the most complete variant of the completely renormalized CR-CC(2,3) approach of Refs. [144, 145], abbreviated sometimes as CR-CC(2,3)$_{\mathrm{D}}$.
[8] Total CCSDTQ energy in hartree.
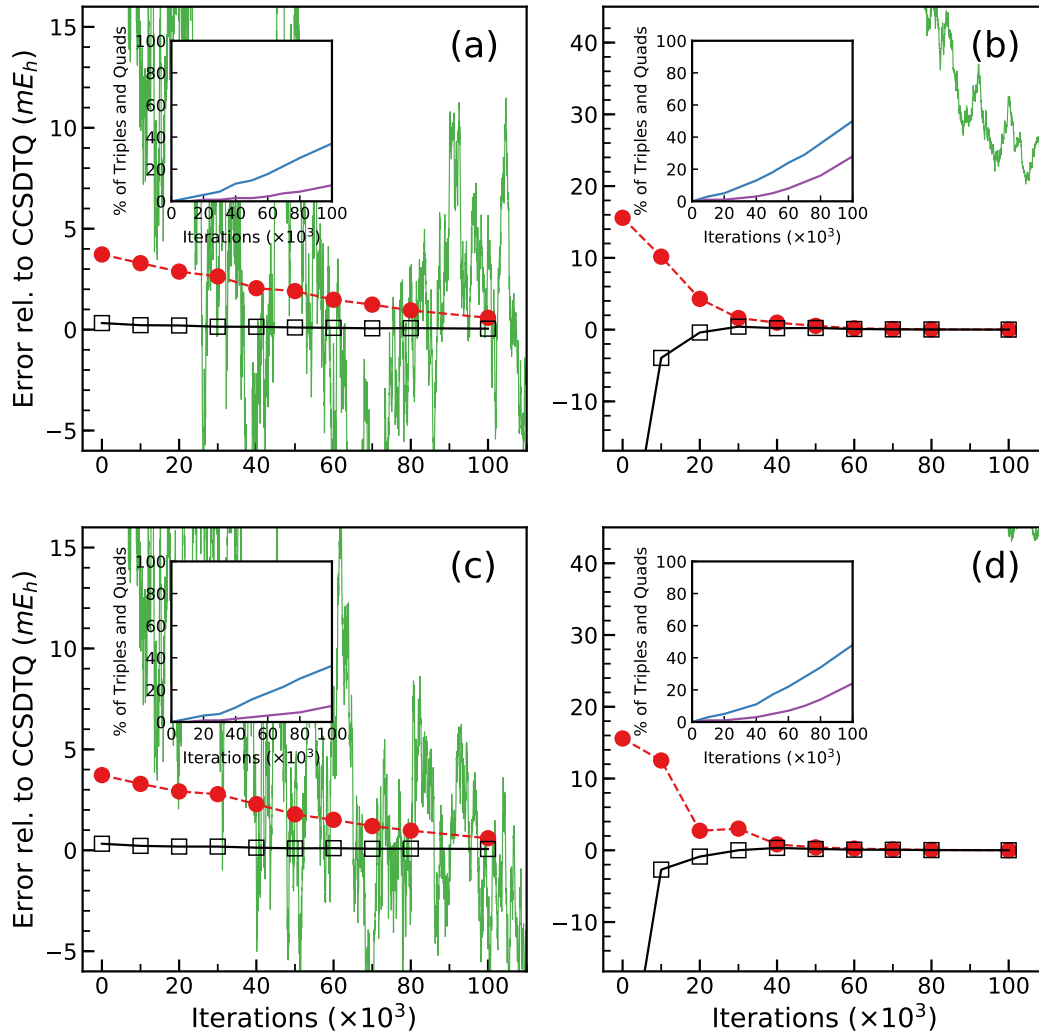
Figure 3.7: Convergence of the CC(P) (red circles and dashed lines) and CC(P;Q)$_{\mathrm{EN}}$ (black open squares and solid lines) energies toward CCSDTQ for the water molecule, as described by the cc-pVDZ basis set. The relevant i-CIQMC runs (all using $\Delta\tau = 0.0001$ a.u.) are depicted by the green lines representing the corresponding projected energies. Panels (a) and (b) correspond to the calculations in which the P spaces employed in the CC(P) steps consisted of all singles and doubles and subsets of triples and quadruples identified during the i-FCIQMC propagations; the Q spaces needed to define the corresponding $\delta(P;Q)$ corrections consisted of the triples that were not captures by the i-FCIQMC. Panels (c) and (d) correspond to the calculations in which the P spaces employed in the CC(P) steps consisted of all singles and doubles and subsets of triples and quadruples identified during the i-CISDTQ-MC propagations; in this case, the Q spaces needed to define the corresponding $\delta(P;Q)$ corrections consisted of the triples that were not captures by i-CISDTQ-MC. Panels (a) and (c) correspond to the equilibrium geometry. Panels (b) and (d) correspond to the geometry in which both O−H bonds in water are simultaneously stretched by a factor of 3 without changing the ∠(H−O−H). All reported energies are errors relative to CCSDTQ in mE_{h}. The insets show the percentages of triples (blue line) and quadruples (purple line) captures during the relevant i-CIQMC propagations.

relative to CCSDTQ, which stand at multiple mE_h below the CCSDTQ energy, to less than $1\,mE_h$ accuracies. This results are evidence that the semi-stochastic CC(P;Q) methods, introduced in this Ph.D. work, are able to correctly recover their parent high-level CC energies at a fraction of the computational cost required for traditional truncated CC methods, showing how flexible and yet powerful the CC(P;Q) hierarchy of methods is in producing qualitatively and quantitatively correct energetics, even when challenging systems are studied. But perhaps more importantly, these results suggest that the semi-stochastic CC(P;Q) methods can be extended to systematically improve CC approximations to any order of CC theory by providing a continuous path to increase the level of theory in-between the traditional truncation schemes (e.g. CCSD, CCSDT, CCSDTQ, etc.).

3.2 Applying the Extension of the Semi-Stochastic CC(P) Approach to Singlet Excited Electronic States of the CH$^+$ Ion

The success of the semi-stochastic CC(P;Q) approach in computing ground-state energies of the challenging bond breaking examples of F_2 and H_2O, and the automerization of cyclobutadiene motivated us to extend our methodology to excited electronic states. We achieved this by employing the EOM framework, in which we diagonalize the similarity transformed, $\bar{H}^{(P)}$, in stochastically sampled subspaces of the Hilbert space, spanned by Slater determinants of spatial symmetries matching the various target excited states of interest.

To test the performance of our semi-stochastic CC(P) methods in computing excited state energies, we computed the first few states of the the CH$^+$ ion at various distances of the C$-$H bond. We studied the ground state, as well as the $2\,^1\Sigma^+$, $3\,^1\Sigma^+$, $4\,^1\Sigma^+$, $1\,^1\Pi$, $2\,^1\Pi$, $1\,^1\Delta$, and $2\,^1\Delta$ states, for the equilibrium geometry, $R_{C-H} = R_e = 2.137\,13\,\text{bohr}$, in addition to the non-equilibrium geometries where $R_{C-H} = 1.5R_e$ and $R_{C-H} = 2R_e$. The

CH$^+$ ion is employed extensively [88–90, 129, 132–134, 139, 140, 146, 149, 150, 162, 250–255] as a test case system for the evaluation of the performance of QC methods capable of describing excited electronic states. The relevance of this particular ion lies in the fact that, most of its excited states, particularly at the stretched geometries, are characterized by considerable multireference correlations dominated by two-electron excitation contributions. This kind of correlation effects is challenging for many QC methods, including the entry-level EOMCCSD method, which produces large errors when most excited states of the CH$^+$ ion are examined. For example, at the equilibrium geometry, EOMCCSD energies for the $^1\Delta$ states already underestimate the EOMCCSDT vertical excitation energy by about 35 mE_h. If we consider the geometry where the internuclear separation is set at $2R_e$, these errors climb to about 45–145 mE_h. These huge inaccuracies show the inadequecy of EOMCCSD in describing the CH$^+$ ion, and highlight how useful this molecular system is in assessing the performance of QC methods.

In the next subsections, we will present numerical results that demonstrate the capability of the semi-stochastic CC(P) methodology in capturing most of the aforementioned correlation effects. We will show how the huge errors of EOMCCSD quickly vanish by including triply excited cluster and linear excitation components to $T^{(P)}$ and $R^{(P)}$ from lists of Slater determinants obtained during i-CIQMC simulations.

3.2.1 Computational Details

As was the case with the previous examples presented in this work, the integrals defining the one- and two-body terms of the Hamiltonian were computed using molecula orbitals of RHF calculations performed in the GAMESS (US) software package. Since CH$^+$ belongs to the $C_{\infty v}$ point-group symmetry, we ran the RHF calculations enforcing C_{2v} point-group symmetry, which is a subgroup of $C_{\infty v}$ and it is also abelian. The lists of determinants required to build the appropriate P space for the ground-state calculations, where the cluster operator is approximated by $T^{(P)} = T_1 + T_2 + T_3^{(\mathrm{MC})}$, were obtained

from *i*-FCIQMC simulations within the totally symmetric A_1 irreducible representation. Once the ground state was calculated, and the similarity transformed Hamiltonian, $\bar{H}^{(P)}$, generated, the EOMCC(P) vertical excitation energies were computed by diagonalizing $\bar{H}^{(P)}$ in the P space within the irreducible representations of C_{2v} corresponding to the excited states of interest. The various P spaces were obtained from *i*-FCIQMC simulations carried out within the $^1A_2$, $^1B_1$, and $^1B_2$ irreducible representations resulting in the lowest energy states of each symmetry. For example, all $^1\Sigma$ states were computed using the same $^1A_1$ P space previously employed in in the ground state. The $^1\Pi$ states were calculated using P spaces within the $^1B_1$ irreducible representations, and the $^1\Delta$, with the P spaces within the $^1A_2$ irreducible representations. All *i*-FCIQMC propagations were carried out using the HANDE QMC software [243, 244] by setting the initiator threshold parameters, $n_a = 3$, $\Delta\tau = 0.0001$ a.u., and by placing 1500 walkers on the reference determinant (the ground-state RHF determinant for the $^1\Sigma^+$ states, $3\sigma \rightarrow 1\pi$ state of the $^1B_1(C_{2v})$ symmetry for the $^1\Pi$ states, and $3\sigma^2 \rightarrow 1\pi^2$ state of the $^1A_2(C_{2v})$ symmetry for the $^1\Delta$ states). The CC(P) and EOMCC(P) calculations were performed using our group's own codes.

3.2.2 Numerical Results

As discussed above, describing the excited electronic states of the CH^+ molecule accurately is quite a challenging feat, mainly due to its strong multireference character in which electronic correlations are dominated by two-electron excitations. This issue renders many traditional QC approximated methods useless, including the EOMCCSD approach. At the equilibrium geometry, our results, shown in Table 3.7, improve upon the failure of EOMCCSD, which produces energies with huge, 20–35 mE_h, errors relative to EOMCCSDT, reaching sub-mE_h accuracies during the early stages of the *i*-FCIQMC simulations. At 50 000 iterations of the *i*-FCIQMC simulation subject to the $^1A_1$ irrep, 37 % of triples are recovered, which translates in accuracies below 1 mE_h for all the $^1\Sigma$

states. At this stage, the most challenging $2\,^1\Sigma^+$ state, which at the EOMCCSD level of theory is under correlated by about $25\,\mathrm{m}E_\mathrm{h}$ compared to EOMCCSDT, reaches an error of only $0.944\,\mathrm{m}E_\mathrm{h}$. The ground state is able to arrive at similar accuracies much earlier, at only 4000 iterations and with only 11 % of the relevant triples. The remaining EOMCC(P) calculations of the $3\,^1\Sigma^+$ and $4\,^1\Sigma^+$, which share the P space of the previous two states and the ground state, arrive at equivalent accuracies at 8000 iterations of the i-FCIQMC simulation with only 23 % of the triples.

A similar story can be told about the EOMCC(P) calculations for the $^1\Pi$ and $^1\Delta$ states. By 6000 iterations and 25 %, the energy of both $^1\Pi$ states is calculated with errors below $1\,\mathrm{m}E_\mathrm{h}$. But more interestingly, the massive errors of 34.304 and $34.685\,\mathrm{m}E_\mathrm{h}$ belonging to the EOMCCSD energies of the $1\,^1\Delta$ and $2\,^1\Delta$ states, are quickly reduced to only 0.240 and $0.940\,\mathrm{m}E_\mathrm{h}$, respectively, when the 8000 iterations mark is reached by the i-FCIQMC simulation, by which time 22 % of the $^1A_2$ symmetric triples are captured.

The remarkable performance trend of the EOMCC(P) method is further supported by the results of the examination of the $R = 1.5R_e$ geometry, shown in Table 3.8. As in the case of $R = R_e$, by 50 000 iterations and 32–37 % of the corresponding triples, the EOMCC(P) energies are already at the sub-mE_h level. This is a remarkable success, since the most challenging states, represented by $2\,^1\Sigma^+$, $2\,^1\Pi$, and both $1\,^1\Delta$ and $2\,^1\Delta$, produce huge errors on the order of 21.380–79.183 $\mathrm{m}E_\mathrm{h}$ relative to EOMCCSDT energies, at the EOMCCSD level of theory. If we disregard the $2\,^1\Delta$ state, which has the largest inaccuracy at the EOMCCSD level of theory, the EOMCC(P) energies of the remaining $2\,^1\Sigma^+$, $2\,^1\Pi$, and $1\,^1\Delta$ states achieve sub-millihartree accuracy earlier, with $2\,^1\Sigma^+$ reaching it at 10 000 iterations and 19 % triples, $2\,^1\Pi$, at 6000 with 16 % triples, and $1\,^1\Delta$ 4000 with 12 % triples. By 200 000 iterations, all vertical excitation energies are around $0.01\,\mathrm{m}E_\mathrm{h}$.

Finally, the even more complicated energies of the $R = 2R_e$ geometry, where the error of EOMCCSD energies calculated for the $2\,^1\Delta$ state relative to the EOMCCSDT energy is as high as $144.414\,\mathrm{m}E_\mathrm{h}$, tells the same story as the above-discussed cases. Following the

Table 3.7: Convergence of the CC(P) energies toward the CCSDT and EOMCC(P) energies of the three lowest-energy excited states of the $^1\Sigma^+$ symmetry, two lowest states of $^1\Pi$ symmetry, and two lowest $^1\Delta$ states toward EOMCCSDT for the CH$^+$ ion, as described by the [5s3p1d/3s1p] basis set,[d] at equilibrium internuclear separation $R = R_e = 2.137\,13$ bohr. The P space used in the CC(P) and EOMCC(P) calculations for the $^1\Sigma^+$ states consisted of all singles and doubles and subsets of triples extracted from the i-FCIQMCpropagation. The P spaces used in the EOMCC(P) diagonalizations for the $^1\Pi$ and $^1\Delta$ states consisted of all singles and doubles and subsets of triples extracted from the i-FCIQMCpropagations for the lowest-energy states of the relevant symmetries. Each i-FCIQMC calculation preceding the CC(P) and EOMCC(P) steps was initiated by placing 1500 walkers on the corresponding reference function (the ground-state RHF determinant for the $^1\Sigma^+$ states, $3\sigma \to 1\pi$ state of the $1B(C_{2v})$ symmetry for the $^1\Pi$ states, and the $3\sigma^2 \to 1\pi^2$ state of the $^1A_2(C_{2v})$ symmetry for the $^1\Delta$ states). The n_a parameter of the initiator algorithm was set at 3 and the time step $\Delta\tau$ used in the i-FCIQMC run was 0.0001 a.u.

MC iter. (10^3)	1 $^1\Sigma^+$ ΔE[a]	%T[b]	2 $^1\Sigma^+$ ΔE[a]	3 $^1\Sigma^+$ ΔE[a]	4 $^1\Sigma^+$ ΔE[a]	1 $^1\Pi$ ΔE[a]	%T[b]	2 $^1\Pi$ ΔE[a]	1 $^1\Delta$ ΔE[a]	%T[b]	2 $^1\Delta$ ΔE[a]
0[c]	1.845	0	19.694	3.856	5.537	3.080	0	11.656	34.304	0	34.685
2	1.071	7	11.004	3.248	4.826	0.772	13	3.746	1.492	10	5.951
4	0.423	15	5.474	1.893	1.980	0.513	20	1.852	0.525	16	2.542
6	0.249	20	4.712	1.268	1.077	0.213	25	0.957	0.471	18	1.892
8	0.181	23	1.371	0.643	0.702	0.170	27	0.743	0.240	22	0.940
10	0.172	24	1.572	0.295	0.385	0.118	29	0.411	0.198	24	0.877
50	0.077	37	0.755	0.139	0.208	0.053	43	0.157	0.039	42	0.133
100	0.044	48	0.277	0.007	0.155	0.021	57	0.063	0.014	56	0.043
150	0.015	59	0.085	0.058	0.041	0.008	71	0.020	0.004	71	0.008
200	0.006	69	0.024	0.014	0.002	0.004	82	0.008	0.003	82	0.003

[a] The ΔE values are errors relative to the CCSDT and EOMCCSDT energies, in mE_h, obtained in Refs.. The CCSDT energy for the $^1\Sigma^+$ state is $-38.019\,516\,E_h$. The EOMCCSDT energies 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 are $-37.702\,621\,E_h$, $-37.522\,457\,E_h$, $-37.386\,872\,E_h$, $-37.900\,921\,E_h$, $-37.498\,143\,E_h$, $-37.762\,113\,E_h$, and $-37.402\,308\,E_h$, respectively.

[b] The %T values are the percentages of triples captured during the i-FCIQMCpropagations for the lowest-energy state of a given symmetry [the 1 $^1\Sigma^+$ = 1 $^1A_1(C_{2v})$ ground state for the $^1\Sigma^+$ states, the $B_1(C_{2v})$ component of the 1 $^1\Pi$ state for the $^1\Pi$ states, adn the $A_2(C_{2v})$ component of the 1 $^1\Delta$ state for the $^1\Delta$ states].

[c] The CC(P) and EOMCC(P) energies at $\tau = 0$ a.u. are identical to the energies obtained in the CCSD and EOMCCSD calculations.

[d] Basis set described in Ref. [255].

Table 3.8: Same as Table 3.7 for the stretched internuclear separation $R = 1.5R_e = 3.205\,695$ bohr[a].

MC iter. (10^3)	1 $^1\Sigma^+$ ΔE[a]	%T[b]	2 $^1\Sigma^+$ ΔE[a]	3 $^1\Sigma^+$ ΔE[a]	4 $^1\Sigma^+$ ΔE[a]	1 $^1\Pi$ ΔE[a]	%T[b]	2 $^1\Pi$ ΔE[a]	1 $^1\Delta$ ΔE[a]	%T[b]	2 $^1\Delta$ ΔE[a]
0	2.815	0	25.344	6.513	10.885	6.769	0	21.380	41.207	0	79.183
2	1.329	4	14.788	2.731	9.417	2.182	10	8.773	1.298	8	7.161
4	0.645	11	5.850	1.597	6.031	1.035	16	4.212	0.534	12	4.002
6	0.321	15	2.489	0.501	1.926	0.447	18	0.557	0.305	15	1.812
8	0.196	17	1.132	0.279	1.122	0.523	21	0.713	0.303	18	1.298
10	0.176	19	0.944	0.325	2.536	0.198	23	0.430	0.209	20	1.007
50	0.087	32	0.399	0.214	0.190	0.037	37	0.097	0.050	33	0.153
100	0.018	44	0.114	0.075	0.118	0.008	49	0.015	0.004	46	0.014
150	0.010	54	0.095	0.056	0.028	0.001	61	0.008	0.002	59	-0.006
200	0.003	65	0.011	0.007	0.008	0.001	73	0.006	0.001	71	0.001

[a] The ΔE values are errors relative to the CCSDT and EOMCCSDT energies, in mE_h, obtained in Refs.. The CCSDT energy for the $^1\Sigma^+$ state is $-37.954\,008\,E_h$. The EOMCCSDT energies 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 are $-37.696\,428\,E_h$, $-37.609\,784\,E_h$, $-37.438\,611\,E_h$, $-37.890\,831\,E_h$, $-37.650\,437\,E_h$, $-37.736\,600\,E_h$, and $-37.440\,775\,E_h$, respectively.

Table 3.9:   Same as Table 3.7 for the stretched internuclear separation $R = 2R_e = 4.274\,26\,\text{bohr}$[a].

MC iter. (10³)	1 ¹Σ⁺		2 ¹Σ⁺	3 ¹Σ⁺	4 ¹Σ⁺	1 ¹Π		2 ¹Π	1 ¹Δ		2 ¹Δ
	ΔE [a]	%T[b]	ΔE [a]	ΔE[a]	ΔE[a]	ΔE[a]	%T[b]	ΔE[a]	ΔE[a]	%T[b]	ΔE[a]
0	5.002	0	17.140	19.929	32.639	13.552	0	21.200	44.495	0	144.414
2	1.588	3	5.209	12.524	33.400	1.398	7	1.644	1.372	6	13.363
4	0.504	7	4.263	6.383	12.671	0.712	12	0.724	0.451	9	3.338
6	0.275	11	1.405	1.352	5.870	0.409	14	0.612	0.422	12	2.340
8	0.263	12	1.543	1.173	4.406	0.436	16	0.457	0.253	13	2.088
10	0.148	14	0.792	0.613	2.331	0.227	17	0.220	0.122	14	0.862
50	0.030	26	0.302	0.339	0.457	0.061	30	0.079	0.047	26	0.288
100	0.009	39	0.103	0.119	0.110	0.013	41	0.016	0.013	36	0.038
150	0.004	52	0.031	0.035	0.076	0.005	52	0.007	0.005	47	0.014
200	0.001	63	0.024	0.019	−0.006	0.002	65	0.001	0.001	57	0.003

[a] The ΔE values are errors relative to the CCSDT and EOMCCSDT energies, in mE_h, obtained in Refs.. The CCSDT energy for the ¹Σ⁺ state is $-37.900\,394\,E_h$. The EOMCCSDT energies for the 2 ¹Σ⁺, 3 ¹Σ⁺, 4 ¹Σ⁺, 1 ¹Π, 2 ¹Π, 1 ¹Δ, and 2 ¹Δ states are $-37.704\,834\,E_h$, $-37.650\,242\,E_h$, $-37.495\,275\,E_h$, $-37.879\,532\,E_h$, $-37.702\,345\,E_h$, $-37.714\,180\,E_h$, $-37.494\,031\,E_h$, respectively.

previous examples, and as shown in Table 3.9, sub-mE_h accuracies are achieved with a small fraction of the relevant triples, in this case with 26–30 % triples which correspond to 50 000 iterations of the relevant *i*-FCIQMC simulation. With the exception of the $4\,^1\Sigma^+$ state, this sub-mE_h accuracy is reached for the remaining states at a much earlier stage, namely at 10 000, which corresponds to about 14–17 % triples of the appropriate symmetry. This fact is particularly impressive when examining the $2\,^1\Delta$ state, since the massive error of 144.414 mE_h from EOMCCSD calculations, relative to EOMCCSDT, becomes a mere 0.862 mE_h with 14 % of the $^1A_2$ symmetric triples, which represents a remarkable improvement.

In summary, our first exploration of semi-stochastic CC approaches exploiting the *i*-CIQMC methods, shows that indeed, they provide a straight forward manner for systematically and efficiently improving CC methods to account for high-level correlation effects both in the ground and exited states. Our semi-stochastic EOMCC(P) approach shows that higher order CC methods including higher-than-doubly excited components, which are usually sought when dealing with multireference situations such as the ones described in this section, are not required to be treated fully to achieve high-level accuracies, resulting in considerable computational savings.

3.3 Approaching the Exact Quantum Chemistry via Cluster Analysis of the FCIQMC Wave Function

Another useful way of taking advantage of stochastic methods such as *i*-FCIQMC, is to utilize their resulting wave function directly. As mentioned in the previous chapter, this can be achieved by the historically recognized cluster analysis of the FCI wave function, where CI coefficients are translated into cluster amplitudes, which can then be employed in the solution of CC-like equations, as in the so-called externally corrected methods. The CAD-FCIQMC method, introduced during this Ph.D. dissertation in collaboration with my colleague Ilias Magoulas, exploits this fact, by extracting the *i*-FCIQMC-determined C_1 through C_4 CI coefficients, translating them into T_3 and T_4 cluster amplitudes, which are subsequently used to attain high-quality energetics of the exact, FCI, type, whose accuracy uniquely depend on the quality of the input wave function used during the cluster analysis.

In the first section of the current chapter, we explained that the example given by the simultaneous bond stretching of the H_2O molecule represent a challenging system to most QC methods. We discussed how CCSD, CCSDT, and even CCSDTQ energies, although rapidly converging to the FCI limit, fail to perfectly reproduce FCI energetics, especially when stretched geometries are examined. This means that the H_2O molecule, again, represents a great example for testing the performance of the CAD-FCIQMC method, since the goal is to recover the exact FCI energetics.

Of course, in the CAD-FCIQMC approach, we are not solving fully deterministic CC equations as opposed to the semi-stochastic CC(P;Q) method introduced in this work. The CAD-FCIQMC method takes advantage of coupling T_1 and T_2 cluster amplitudes with $T_3^{(\text{MC})}$ and $T_4^{(\text{MC})}$, which are acquired from *i*-FCIQMC simulations. Since the CC energy is only dependent on T_1 and T_2 and, in the CAD-FCIQMC approach, T_1 and T_2 are coupled with $T_3^{(\text{MC})}$ and $T_4^{(\text{MC})}$, the energies produced by CAD-FCIQMC are highly dependent on the quality of the wave function being cluster analyzed. On one hand, this

can be an issue, since the CAD-FCIQMC energies are subject to the noise of the stochastic simulations. However, since the i-FCIQMC approach is guaranteed to converge toward the true FCI energy, provided that long simulations are conducted, the CAD-FCIQMC will too. In the worst case, when the quality of the T_3^{MC} and $T_4^{(MC)}$ cluster components is not very good, due to the underlying i-FCIQMC simulation being in its early stages, the CAD-FCIQMC method will filter out the extra noise produced by i-FCIQMC eventually converging to the exact FCI energy.

3.3.1 Computational Details

In order to test the performance of the CAD-FCIQMC method, we examined the equilibrium geometry, where $R_{O-H} = R_e$, and two stretched geometries corresponding to 1.5 and 2 times R_e. All one- and two-body integrals used in the computation were obtained from the GAMESS (US) software suite using the cc-pVDZ basis set as in the previous example. The i-FCIQMC simulations were carried out using HANDE [243, 244] by placing 1000 walkers on the RHF determinant, setting the time step, $\Delta\tau$, to 0.0001 a.u., and adjusting the initiator threshold parameter n_a to the default value of 3. The i-FCIQMC wave functions, extracted from the imaginary time propagation every 1000 iterations, were cluster analyzed by our own codes and subsequently, the resulting $T_1^{(MC)}$, $T_2^{(MC)}$, $T_3^{(MC)}$, and $T_4^{(MC)}$ cluster analyzed amplitudes were introduced in CC codes. In these CCSD-like calculations, both $T_3^{(MC)}$ and $T_4^{(MC)}$ amplitudes were kept frozen while the $T_1^{(MC)}$ and $T_2^{(MC)}$ were allowed to relax. This procedure was repeated every 1000 iterations during the first 160 000 iterations of the corresponding i-FCIQMC propagations.

3.3.2 Numerical Results

As can be seen from the results shown in Tables 3.10 and 3.11 and in Figure 3.8, the CAD-FCIQMC results converge much smoother and relatively faster than the i-FCIQMC projective energy values as a function of the imaginary time. The CAD-FCIQMC propa-

gations of the R_e, $1.5R_e$, and $2R_e$ geometries, have the advantage that their starting point at 0 iterations, corresponds to the CCSD ground-state energy. This is, of course, already a huge improvement upon the RHF energy, which characterizes the beginning of all i-FCIQMC simulations. However, the CAD-FCIQMC energies are able to rapidly improve their CCSD starting point, achieving FCI-quality energetics with high accuracies within the sub-mE_h level. At the equilibrium geometry, CAD-FCIQMC energies achieve stable, lower than $1\,\mathrm{m}E_h$, accuracies after only $10\,000$ iterations, in which i-FCIQMC simulation capture about $10\,\%$ of triples and $1\,\%$ of quadruples. Similarly, at the stretched geometries, where $R_{\mathrm{O-H}} = 1.5R_e$ and $R_{\mathrm{O-H}} = 2R_e$, the corresponding CAD-FCIQMC results fall below the $1\,\mathrm{m}E_h$ mark, quickly, after only $60\,000$ and $100\,000$ iterations, respectively, by capturing $\sim 25\,\%$ triples and $\sim 5\,\%$ quadruples, and $\sim 40\,\%$ triples $\sim 10\,\%$ quadruples of the corresponding Hilbert spaces. This is an interesting result, since the errors of CCSD energetics relative to FCI are quite large, especially at the stretched geometries of the H_2O molecule, and the fractions of triples and quadruples needed to achieve sub-mE_h results is impressively low, even at the more complicated $R_{\mathrm{O-H}} = 2R_e$ geometry.

The stochastic nature of the amplitudes in $T_3^{(\mathrm{MC})}$ and $T_4^{(\mathrm{MC})}$ cluster components entering our CAD-FCIQMC calculations, indicate the need for time-series statistical analysis tools that correct for the correlation between time-correlated steps, such as the re-blocking analysis [228]. However, since our raw results converge so fast toward the right FCI-limit, while the i-FCIQMC are still in the equilibration phase, and our goal is to introduce a novel idea on how to exploit i-FCIQMC-generated wave functions, we did not run our simulations long enough to collect the necessary data to perform such analysis. Nevertheless, we compiled the convergence information in Table 3.11, where we compare the mean and standard deviations of our CAD-FCIQMC and i-FCIQMC simulations as a function of imaginary time, and where one can clearly see the advantage in the overall numerical stability introduced by CAD-FCIQMC calculations.

Our CAD-FCIQMC are only raw results showing the potential of exploiting the i-
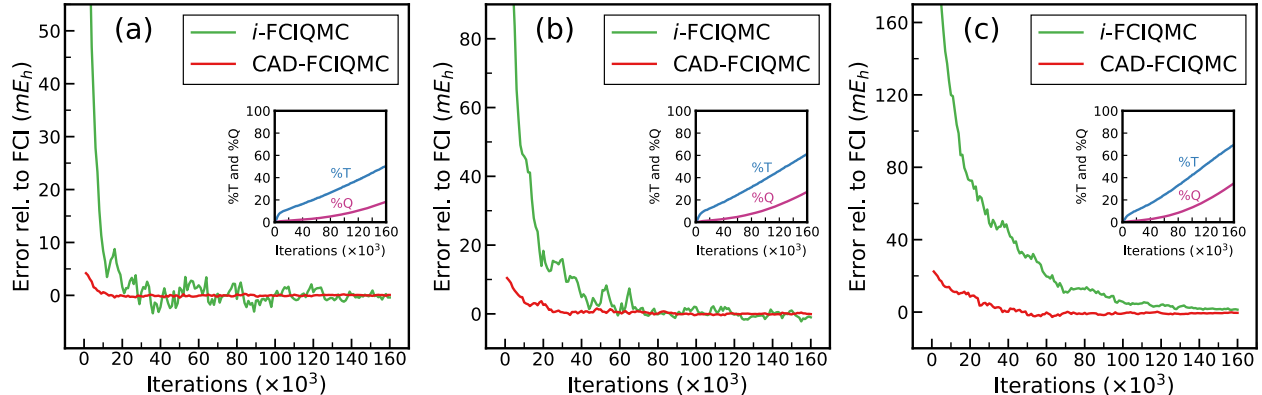
Figure 3.8: Convergence of the energies resulting from the all-electron *i*-FCIQMC and CAD-FCIQMC calculations $\Delta\tau = 0.0001$ a.u. toward FCI at equilibrium and two displaced geometries corresponding to a simultaneous stretching of both O−H bonds in the H_2O/cc-pVDZ molecule by factors of 1.5 and 2. Panel (a) corresponds to the equilibrium geometry. Panel (b) and (c) correspond to the geometries that represent a simultaneous stretching of both O−H bonds by factors of 1.5 and 2.0, respectively, without changing the $\angle$(H−O−H) angle. The *i*-FCIQMC calculations were initiated by placing 1000 walkers on the RHF determinant and the n_a parameter of the initiator algorithm [1] was set at the default value of 3. All energies are errors relative to FCI in mE_h, and the insets show the percentages of triply (%T) and quadruply (%Q) excited determinants captured during the *i*-FCIQMC propagations.

FCIQMC wave function in the framework of the externally corrected CC methods. Therefore, further investigation is required to show other promising advantages of the CAD-FCIQMC method, including the use of the reblocking analysis. With this in mind, we envision the utility of the CAD-FCIQMC approach to be in the acceleration of *i*-FCIQMC calculations themselves. For example, one could take advantage of the numerical values of T_1 and T_2 clusters obtained from solving CAD-FCIQMC equations, by translating them to walker populations in *i*-FCIQMC simulations, such that C_n coefficients in the stochastic wave function are improved during the imaginary-time propagation.

Table 3.10: Convergence of the energies resulting from the all-electron i-FCIQMC and CAD-FCIQMC calculations with $\Delta\tau = 0.0001$ a.u. toward FCI at the equilibrium and two displaced geometries corresponding to a simultaneous stretching of both O−H bonds in the H_2O/cc-pVDZ molecule by factors of 1.4 and 2.[1] The i-FCIQMC calculations were initiated by placing 1000 walkers on the RHF determinant and the n_a parameter of the initiator algorithm was set at 3.

Iterations	$R_{O-H} = R_e$		$R_{O-H} = 1.5R_e$		$R_{O-H} = 2R_e$	
	CAD-FCIQMC	i-FCIQMC	CAD-FCIQMC	i-FCIQMC	CAD-FCIQMC	i-FCIQMC
0	3.744[2]	217.821[3]	10.043[2]	269.961[3]	22.032[2]	363.954[3]
10000	0.611	8.381	3.335	45.802	12.351	119.896
20000	−0.073	1.596	2.597	18.345	8.485	72.650
30000	−0.175	0.586	0.473	15.937	4.794	49.203
40000	−0.211	−2.217	0.873	4.855	0.138	44.627
50000	−0.440	1.456	1.310	3.247	−1.693	31.448
60000	−0.046	1.911	0.501	0.588	−0.225	19.660
70000	−0.235	0.302	0.685	0.811	−0.377	11.333
80000	0.189	−0.686	0.063	−0.128	−0.425	12.611
90000	−0.177	−0.981	−0.171	0.014	−1.657	10.089
100000	−0.036	0.139	−0.302	1.956	−0.816	5.680
110000	0.129	−0.710	−0.189	1.088	−0.580	4.797
120000	−0.035	0.597	0.020	−0.241	−0.555	4.041
130000	0.086	−0.503	−0.020	−0.720	−1.166	3.107
140000	0.098	0.080	−0.084	0.497	−0.666	1.981
150000	−0.055	0.308	−0.156	0.990	−0.620	1.630
160000	0.078	−0.400	−0.059	−1.002	−0.434	1.328
∞	-76.241860[4]		-76.072348[4]		-74.951665[4]	

[1] The equilibrium geometry, $R_{O-H} = R_e$, and the geometry that represent a simultaneous stretching of both O−H bonds by factors of 1.5 and 2.0 without changing the ∠(H-O-H) angle were taken from Ref. [242]. Unless otherwise stated, all energies are errors relative to FCI in millihartree.
[2] Equivalent to CCSD.
[3] Equivalent to RHF.
[4] Total FCI energy taken from Ref. [242].

Table 3.11: Statistical analysis of the *i*-FCIQMC and CAD-FCIQMC calculations for the water molecule shown in Table 3.10 and Figure 3.8.

Iterations (I)	%T[1]	%Q[2]	f_I[3]	m_I[4] CAD-FCIQMC	m_I[4] *i*-FCIQMC	σ_I[5] CAD-FCIQMC	σ_I[5] *i*-FCIQMC
				$R_{\text{O-H}} = R_e$			
0	0	0	0.09	0.087	4.035	0.593	18.944
20,000	11.8	1.38	3.95	−0.036	0.109	0.128	1.471
40,000	16.4	2.28	6.52	−0.020	0.043	0.122	1.298
60,000	21.3	3.48	10.3	0.003	−0.004	0.101	1.171
80,000	26.5	5.12	16.1	0.013	−0.165	0.095	1.043
100,000	32.1	7.47	25.7	0.017	0.055	0.080	0.823
120,000	38.0	10.3	40.0	0.049	−0.099	0.053	0.716
140,000	44.2	14.0	63.5	0.029	−0.075	0.055	0.354
				$R_{\text{O-H}} = 1.5R_e$			
0	0	0	0.03	0.853	10.496	1.816	29.168
20,000	12.4	1.58	1.72	0.280	2.528	0.476	4.511
40,000	17.9	2.85	3.33	0.202	0.914	0.385	1.976
60,000	24.7	4.77	6.15	0.091	0.360	0.272	1.355
80,000	31.5	7.46	11.1	0.019	0.067	0.215	0.996
100,000	38.6	11.1	19.6	−0.019	−0.081	0.155	0.983
120,000	46.1	15.7	34.2	0.021	−0.527	0.139	0.744
140,000	53.6	21.1	58.7	0.062	−0.346	0.168	0.799
				$R_{\text{O-H}} = 2R_e$			
0	0	0	0.01	1.645	32.718	5.095	50.838
20,000	10.6	1.34	0.55	−0.063	17.098	2.035	18.864
40,000	16.6	2.72	1.29	−0.721	10.707	0.730	10.739
60,000	24.5	5.03	2.85	−0.776	6.492	0.468	4.981
80,000	33.2	8.46	6.01	−0.709	4.574	0.356	3.268
100,000	42.5	13.2	12.4	−0.619	2.951	0.298	1.285
120,000	51.7	19.2	25.1	−0.711	2.241	0.253	0.841
140,000	60.9	26.6	50.3	−0.529	1.651	0.154	0.229

[1] Percentages of triply excited determinants captured during *i*-FCIQMC propagations.

[2] Percentages of quadruply excited determinants captured during *i*-FCIQMC propagations.

[3] Walker populations characterizing the *i*-FCIQMC propagations reported as percentages of the total walker numbers at $I = I_{\max} = 160\,000$, which in specific *i*-FCIQMC runs shown in Table 3.10 and Figure 3.8 were $1\,169\,396$ at $R_{\text{O−H}} = R_e$, $3\,327\,438$ at $R_{\text{O−H}} = 1.5R_e$, and $10\,146\,724$ at $R_{\text{O−H}} = 2R_e$.

[4] The mean signed error (m_I) and standard deviation (σ_I) values relative to FCI, in mE_{h}, calculated using subsets of either the *i*-FCIQMC or CAD-FCIQMC energies that were obtained at the MC iterations, I, $I + \Delta I$, $I + 2\Delta I, \ldots, I_{\max}$, where $I_{\max} = 160\,000$ is the final iteration at which the *i*-FCIQMC propagations were stopped and ΔI is an iteration increment used to select the energy points for statistical analysis. In the case of the calculations reported once every 20 MC iterations, the FCIQMC-corrected CCSD calculations were performed only once every 1000 iterations). The $m_{I=0}$ and $\sigma_{I=0}$ values correspond to averaging the entire energy dataset characterizing a given method and geometry. The m_I and σ_I values for $I > 0$ correspond to averaging a subset of energy points that starts at the MC iteration I and ends up at the last iteration $I_{\max}$.

[5] The initial walker population, meaning 1000 walkers placed on the RHF reference determinant.

CHAPTER 4

COMPUTATIONAL IMPLEMENTATION AND ALGORITHMIC DETAILS OF THE SEMI-STOCHASTIC CC(P;Q) AND CAD-FCIQMC APPROACHES

In the previous chapters, we have introduced the standard SRCC and EOMCC theories, their active-space extension, and the CC(P;Q) formalism, leading to the semi-stochastic and CAD-FCIQMC methods newly developed in this Ph.D. work. We have also shown some initial results showing the great performance of these new approaches when exploring challenging problems in QC, such as bond breaking events and excited electronic states dominated by two-electron transitions.

Thus far, we have presented and explained the CC and EOMCC ansätze, as well as the corresponding energy corrections within the CC(P;Q) formalism, using operator algebra expressions, which abstract the matrix operators, and their internal contractions, which are required in the computer implementations of these theories. In particular, and despite the fact that we have claimed computer savings associated with the use of our semi-stochastic CC(P;Q) and CAD-FCIQMC calculations, we have not descended into the details about our computer algorithms and methods, which take advantage of the reduced size of the P and Q spaces and FCI wave functions resulting from i-FCIQMC simulations. In the following sections, we will address this by explaining the computational implementation and algorithmic details of both methodologies. But before we dive into them, let us first discuss some fundamentals concerning diagrammatic methods, briefly, by reviewing some key elements of its techniques, which will be used throughout this chapter (cf. [256] for a detailed video-lecture introduction into diagrammatic techniques in QC, which can be followed using the notes in Ref. [257]).
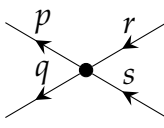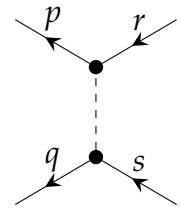
In quantum field theory, diagrams are powerful tools that allows us to derive and organize algebraic expressions resulting from Wick's theorem. This theorem, which is central to any many-body theory, allows us to reduce high-order derivatives, arising from

the quantization of fields, to a combinatorics problems. The diagrams, thus, provide a graphical representation of combinatorial relations between second-quantized operators, thereby simplifying the laborious task of directly manipulating the expressions resulting from the anticommutation algebras defining Wick's theorem. Originally introduced by Feynman, the use of diagrams was intended to aid in calculations within the time-dependent formalism of quantum field theory. Čížek and Paldus, however, proposed that the same techniques could be released from their time constraints [25, 26, 258], and, thus, be applied to time-independent formulations of the Schrödinger equation, which is central to many-body methods in QC (cf. [256, 257, 259–263] for a thorough introduction into diagrammatic techniques).

There are two main components defining time-independent Hugenholtz and Brandow diagrams used in CC theory. The first class are arrows carrying indices, which are portrayed along the x-axis corresponding to the time direction, and which are called fermion lines. These objects directly map to the relevant creation and annihilation operators characterizing second-quantized formulations of many-body operators in the Hilbert space. The second class consists of dashed or wavy lines, or also rectangles, which are oriented along the y-axis corresponding to the space direction, and which represent the many-body amplitudes entering the various many-body operators. These elements are exploited in slightly differently ways when used in the construction of Hugenholtz and their corresponding Brandow diagrams. In Hugenhotz diagrams, one exploits the antisymmetrized matrix elements of many-body operators, by collapsing the multiple nodes appearing in Brandow diagrams, which correspond to creation/annihilation pairs, into a single node. This provides the advantage, unlike Goldstone diagrams, that equivalent diagrams resulting from the antisymmetric nature of many-fermion operators do not have to be explicitly drawn, significantly reducing the amount of diagrams required to calculate algebraic expressions.

In calculating algebraic expressions, we exploit diagrams by contracting fermion lines

Table 4.1: Hugenhotz and Brandow diagrams, with their corresponding algebraic expression, representing the two-body component of the Hamiltonian operator. The uppercase letter N indicates that the creation/annihilation operators enclosed within the square brackets are in the normal order form with respect to the Fermi vacuum.

Hugenholtz	Brandow

$$\frac{1}{4} v_{pq}^{rs} \, N[a^p a^q a_s a_r]$$

between the factors of a product of operators in all possible ways, keeping only topologically unique diagrams and removing redundant equivalences. The resulting diagrams, which contain internal and external lines with free or fixed indices, can be interpreted using a small set of rules. In general, the algebraic expression defining any diagram is given by

$$K_A = s_A w_A \sum_{\chi'_A, \chi''_A} d^{(A)}_{\chi'_A, \chi''_A} O_{\chi''_A}, \tag{4.1}$$

where K_A is the resulting expression, $s_a = (-1)^\rho$ is the sign, w_A is the weight associated with the number of equivalent diagrams of K_A, χ'_A is a summation index over internal lines, χ''_A is a summation index over external lines, $d^{(A)}_{\chi'_A, \chi''_A}$ are scalars indexed with internal and external indices, and $O_{\chi''_A}$ is a product of creation/annihilation operators defining the external lines χ''_A.

To explain how a diagram is interpreted it is best to utilize an example. For instance, let us analyze one of the diagrams resulting from $\langle \Phi_{ij}^{ab} | (V_N T_1 T_2)_C | \Phi \rangle$, where V_N denotes the two-body components of the Hamiltonian in the normal ordered form with respect to the Fermi vacuum and its diagrams are given in Table 4.1. In our technique, we draw all nodes corresponding to components of the cluster and Hamiltonian operators, in both Hugenholtz and Brandow diagrams, with the exception of the excitation operator

89

defining the projection onto the bra state. This omission will be discussed later on. Once we draw the diagrams, as shown in Table 4.2, one can read the algebraic expression as follows. First, we read the weight, w_A, from the Hugenholtz diagram, by finding equivalent lines, which if interchanged do not change the topology of the diagram. The number of possible such operations, g_R, which form classes of automorphisms, is used to calculate the weight of the diagram, such that $w_A = \frac{1}{g_R}$. In our example diagram resulting from $\langle \Phi_{ij}^{ab} | (V_N T_1 T_2)_C | \Phi \rangle$, all equivalence classes are of order 1, and thus $w_A = 1$, since no internal or external line is equivalent to any other. Second, we read the sign, $s_A = (-1)^{l+h}$ from the Brandow diagram, by counting the number of loops, l, and internal hole lines, h, (i.e., lines extending to the right holding an occupied index such as m, n, etc.). In our example, this rule will lead to $s_A = (-1)^{1+2} = -1$, because of the presence one loop formed by lines e and m, and two internal holes lines, m and n. The scalar elements $d_{\chi'_A, \chi''_A}^{(A)}$ are interpreted from the Brandow diagram, by reading the indices of the outgoing and incoming lines at each of the vertices of the operators, which correspond to the lower and upper indices of the matrix representation of the operators. For example, to read the matrix elements corresponding to the operator V_N, we first look at the top vertex, which in this case has an outgoing line with an index n. Thus, we write n as the first lower index (i.e., v_n). Next, we look at the bottom vertex, whose outgoing line is given by index m, and, thus, the put this index in the second position of the lower indices (i.e., v_{nm}). Since no more vertices exist, we go back to the top vertex and this time read the incoming line, given by i, which we write in the first upper index position (i.e., v_{nm}^i). We continue to the bottom vertex, reading the line e and writing the index next to the index i (i.e., v_{nm}^{ie}). Since no more fermion lines leave or arrive at the operator V_N, we can continue with the next operator. We repeat the same procedure with operators T_1 and T_2 by first reading the corresponding weights from the Hugenholtz diagram and sign from the Brandow diagram, and finally by looking at the simple vertices belonging to these operators, in Brandow diagrams, from top to bottom, the first time reading the

indices belonging to outgoing lines, and writing them as lower indices of the elements of the operator, and the second time reading the indices of incoming lines, and writing them as upper indices.

Because the diagrams in our example have external lines we must now find the symmetrization operations, which account for the missing, non-equivalent diagrams, that result from the permutation of occupied, or unoccupied, projection indices i, j, a, and b, which obey $i < j < a < b$, as discussed in Chapter 2. In our example, the permutations correspond to $\mathscr{S}_{ij}$ among occupied indices, and $\mathscr{S}^{ab}$ for unoccupied indices. Since, the permutation of indices results in a change of parity in the algebraic expressions, due to the antisymmetric nature of the fermionic creation and annihilation operators, we express the diagram symmetrizers $\mathscr{S}_{ij}$ and $\mathscr{S}^{ab}$ by the corresponding antisymmetrizers $\mathscr{A}^{ij}$ and $\mathscr{A}_{ab}$ in the algebraic expressions, where $\mathscr{A}^{ij} = 1 - (ij)$ and $\mathscr{A}_{ab} = 1 - (ab)$. In general, antisymmetrizer are defined by the by indices $i, j, k, \ldots$ and $a, b, c, \ldots$, or $i_1, \ldots, i_n$ and $a_1, \ldots, a_n$, such that all permutations among the indices, with their corresponding signs, are included in the operator $\mathscr{A}$. For example, the antisymmetrized $\mathscr{A}^{ijk}$, is given by

$$\mathscr{A}^{ijk} = 1 - (ij) - (ik) - (jk) + (ijk) + (ikj). \tag{4.2}$$

If, however, the indices are separated with the symbol "/", the antisymmetrizer is defined as the operator producing all combinations among the groups of indices separated by "/". For instance,

$$\mathscr{A}^{i/jk} = 1 - (ij) - (ik). \tag{4.3}$$

In the remaining of this Chapter, we will also use the notation $\mathscr{A}^{ij\ldots}_{ab\ldots} = \mathscr{A}^{ij\ldots}\mathscr{A}_{ab\ldots}$ in order to simplify the resulting algebraic expressions.

Table 4.2: Hugenhotz and Brandow diagrams, with their corresponding algebraic expression, representing the $\langle \Phi_{ij}^{ab} | (V_N T_1 T_2)_C | \Phi \rangle$ term entering CC equations.

Hugenholtz	Brandow

$$-\mathscr{A}^{ij} \mathscr{A}_{ab} v_{nm}^{ie} t_a^n t_{eb}^{mj}$$

4.1 Taking Advantage of the Sparse Stochastically Determined *P* Spaces to Improve the Performance of the Semi-Stochastic CC(*P;Q*) Method

Standard CC methods consist in approximating the exact, FCI, energy within a basis set by systematically increasing the maximum rank of the cluster component, $T_n = \frac{1}{n!} t_{a_1 \ldots a_n}^{i_1 \ldots i_n} E_{i_1 \ldots i_n}^{a_1 \ldots a_n}$, allowed in the many-body expansion of Eq. (2.25). In this truncation scheme, both occupied and unoccupied spin-orbital sets of $2n$ indices, labeled by $i_1, \ldots, i_n$, and $a_1, \ldots, a_n$, respectively, are equivalent among their own orbital classes, independently of their label, such that the values of any of the indices in $i_1, \ldots, i_n$ represent the same range over all occupied orbitals and the values of any of the indices in $a_1, \ldots, a_n$, the range over all of the unoccupied orbitals. This constitutes a huge advantage for the derivation and implementation of CC equations, since diagrammatic techniques can be readily applied, without the need to consider extra cases arising from non-equivalent indices. For instance, spin-integrated theories are characterized by occupied and unoccupied indices belonging to the alpha- and beta-spin classes, which are no longer equivalent among each other, prompting the need for extra diagrams capable of describing the additional possible contractions. Therefore, the internal contractions appearing in the diagrams resulting from standard SRCC equations can be easily mapped

to highly optimized and vectorizable matrix-matrix multiplication computations, such as those provided by OpenBLAS, Intel MKL, BLIS [264], and other software frameworks, which considerably improve the performance of the computation of the underlying CC calculations.

Sadly, CC methods based on arbitrary truncations of the cluster operators yielding non-standard P spaces, do not enjoy the same benefit. As discussed before, CC(P;Q) methods attempt to exploit the smaller size of the P spaces, which can be spanned by a smaller set of n-tuply excited determinants relative to the full set defined by standard CC approximations. This leads to cases such as the active-space CC methods, among which the CCSDt, CCSDTq, and CCSDtq are relevant representatives, where some indices in a given component of the cluster operator, T_n, are restricted to an active-space defined using molecular orbitals. In these cases, and unlike in standard CC methods, applying the diagrammatic techniques is not as straight forward. The active and inactive indices, which lead to the core, active occupied, active unoccupied, and virtual labels defining the molecular orbital space (see Figure 2.2), and which are required to define the various projections involving active-space cluster components entering CC equations, lead to many more diagrams that take into account the restrictions over indices that are not equivalent among each other [158]. Despite the computational gains associated with the reduction of the number of projections needed to solve active-space CC equations, the software implementation of these methods becomes considerably more cumbersome, since many more cases defining the possible tensor contractions of Wick's theorem need to be accounted for. This issue becomes even more ponderous when considering spin-integrated active-space CC approaches, in which the number of diagrams grows even larger. Regardless, the diagrams resulting from blocked projections of active-space considerations can be still easily mapped onto matrix-matrix multiplication kernels, and, thus, active-space methods still enjoy from the advantages of highly optimized BLAS routines, despite the extra steps required during the derivation of the corresponding CC

equations.

Unfortunately, the semi-stochastic CC(P;Q) methods introduced in this work further complicate this issue. In contrast to active-space CC methods, the P spaces used in semi-stochastic CC(P;Q) approaches do not allow for blocking schemes, where indices defining the projections and cluster components of the CC(P) equations are labeled according to their active/inactive character. This is a problem, since diagrammatic techniques, as used in the standard CC methods, and in active-space CC methods, where once can easily differentiate between equivalent and non-equivalent lines, are no longer useful. Clearly, if the stochastically determined P space cannot be organized in blocks of labeled indices, the mapping between tensor contractions and matrix-matrix multiplications kernels is lost, and, thus, the performance associated with them.

In response to this, one way of dealing with the implementation of semi-stochastic CC(P;Q) equations is by completely redefining how internal contractions are treated. Instead of employing the usual diagrammatic approach, where tensor contractions are performed among groups of repeating single-particle upper and lower indices, as prescribed by the Einstein summation convention, one can take advantage of matrix elements of the similarity transformed Hamiltonian in the N-electron basis, in a similar way to how CI methods employ matrix elements of the Hamiltonian.

To exemplify this idea, let us examine the semi-stochastic CC(P;Q) method with up to triply excited cluster components, where T_1 and T_2 are treated fully but $T_3^{(\mathrm{MC})}$ is determined stochastically.

As explained in Chapter 2, we begin our considerations by defining the CC equations, characterizing the iterative CC(P) part of the semi-stochastic CC(P;Q) approach, and which are given by

$$\langle \Phi_k | \bar{H}_N^{(P)} | \Phi \rangle = \langle \Phi_k | (H_N e^{T^{(P)}})_C | \Phi \rangle = 0, \quad |\Phi_k\rangle \in \mathscr{H}^{(P)}. \tag{4.4}$$

Let us recall that $\bar{H}_N = (H_N e^{T^{(P)}})_C$ is the similarity transformed Hamiltonian and $H_N = H - \langle \Phi | H | \Phi \rangle = F_N + V_N$ is the Hamiltonian operator in the normal ordered

form with respect to the Fermi vacuum, consisting of the Fock, F_N, and two-body, V_N, operators. The determinants, $|\Phi_k\rangle$, comprise of all singly, $|\Phi_i^a\rangle$, and doubly, $|\Phi_{ij}^{ab}\rangle$, excited determinants, as well as those triply excited determinants, $|\Phi_m\rangle$, captured by the i-CIQMC simulation, the combination of which span the P space. The cluster operator is defined in an analogous manner, that is by $T^{(P)} = T_1 + T_2 + T_3^{(MC)}$, where the cluster components T_1 and T_2 are treated fully, and $T_3^{(MC)}$ is defined using the stochastic information, such that only those triply excited components that belong to the P space are included. If we analyze the terms within $\bar{H}_N^{(P)}$, we can quickly observe that the only non-zero terms are those linear in $T_3^{(MC)}$, since the square of $T_3^{(MC)}$, whose $(V_N T_3^2)_C$ diagrams produce at least 8 lines extending to the left, cannot be projected onto lower than quadruply excited determinants. Since we are addressing the CC(P) approach, where up to triply excited determinants span the P space, we can rewrite the LHS of Eq. (4.4), by

$$\langle\Phi_k|[(H_N e^{T_1+T_2})T_3^{(MC)}]_C|\Phi\rangle\,, \tag{4.5}$$

or simply as the commutator,

$$\langle\Phi_k|[(H_N e^{T_1+T_2})_C, T_3^{(MC)}]|\Phi\rangle\,. \tag{4.6}$$

By expanding the commutator and introducing the resolution of identity between the the similarity transformed Hamiltonian and $T_3^{(MC)}$, we obtain,

$$\begin{aligned}
\langle\Phi_k|(H_N e^{T_1+T_2})_C(|\Phi\rangle\langle\Phi| + \sum_\lambda |\Phi_\lambda\rangle\langle\Phi_\lambda|)T_3^{(MC)}|\Phi\rangle \\
- \langle\Phi_k|T_3^{(MC)}(|\Phi\rangle\langle\Phi| + \sum_{\lambda'}|\Phi_{\lambda'}\rangle\langle\Phi_{\lambda'}|)(H_N e^{T_1+T_2})_C|\Phi\rangle\,,
\end{aligned} \tag{4.7}$$

that readily simplifies to the expression

$$\sum_{k'}\langle\Phi_k|(H_N e^{T_1+T_2})_{C,\text{open}}|\Phi_{k'}\rangle\langle\Phi_{k'}|T_3^{(MC)}|\Phi\rangle\,, \tag{4.8}$$

where $|\Phi_{k'}\rangle$ are triply excited determinants spanning the stochastically determined P space and $(H_N e^{T_1+T_2})_{C,\text{open}}$ is the similarity transformed Hamiltonian having external

lines (i.e., the operator resulting from substracting the closed similarity transformed Hamiltonian possessing no external lines, such that $\bar{H}_{N,\text{open}} = \bar{H}_N - \bar{H}_{N,\text{closed}} = \bar{H}_N - \langle\Phi|\bar{H}_N|\Phi\rangle$). The resulting CC equations, using the expression derived above, are given by,

$$\sum_{k'} \langle\Phi_k|\bar{H}_{N,\text{open}}^{(\text{SD})}|\Phi_{k'}\rangle \langle\Phi_{k'}|T_3^{(\text{MC})}|\Phi\rangle = 0, \quad |\Phi_k\rangle, |\Phi_{k'}\rangle \in \mathscr{H}^{(P)}, \qquad (4.9)$$

where we use $\bar{H}_{N,\text{open}}^{(\text{SD})}$ to refer to the similarity transformed Hamiltonian given by $(H_N e^{T_1 + T_2})_{C,\text{open}}$.

This expression allows us to take advantage of the much smaller stochastically determined P space, since the internal contractions involving $T_3^{(\text{MC})}$ are no longer over molecular orbital indices, which cannot be efficiently blocked, but instead over the indices k' that designate only triply excited determinants, $|\Phi_{k'}\rangle$, determined during the i-CIQMC simulation. On the other hand, even though the determinants $|\Phi_k\rangle$ span the whole stochastically determined P space, we will focus on the optimization of the computations where $|\Phi_k\rangle$ are triply excited determinants belonging to the P space only.

In order to compute the matrix-vector multiplication expression in Eq. (4.9), the next step is to calculate the matrix elements of $\bar{H}_{N,\text{open}}^{(\text{SD})}$. Diagrammatically, this is accomplished by considering all combinations of triply excited determinants which result in non-zero values, which are given by

$$\langle\Phi_k|\bar{H}_n^{(\text{SD})}|\Phi_{k'}\rangle \neq 0, \qquad (4.10)$$

where $\bar{H}_n^{(\text{SD})}$ are the n-body components of the similarity transformed Hamiltonian ranging from $n = 1$ to $n = 3$, such that $\bar{H}_{N,\text{open}}^{(\text{SD})} = \bar{H}_1^{(\text{SD})} + \bar{H}_2^{(\text{SD})} + \bar{H}_3^{(\text{SD})}$, and $\langle\Phi_k|$ and $|\Phi_{k'}\rangle$ are triply excited determinants which differ by at most one spin-orbital when $n = 1$, at most two spin-orbitals when $n = 2$, and at most three spin-orbitals when $n = 3$. For example, the diagonal terms, where $|\Phi_k\rangle = |\Phi_{ijk}^{abc}\rangle$ are triply excited determinants,

result in the following expressions:

$$\langle \Phi_{ijk}^{abc} | \bar{H}_1^{(\mathrm{SD})} | \Phi_{ijk}^{abc} \rangle = - \bar{h}_i^i - \bar{h}_j^j - \bar{h}_k^k$$
$$+ \bar{h}_a^a + \bar{h}_b^b + \bar{h}_c^c, \tag{4.11}$$

$$\langle \Phi_{ijk}^{abc} | \bar{H}_2^{(\mathrm{SD})} | \Phi_{ijk}^{abc} \rangle = + \bar{h}_{ij}^{ij} + \bar{h}_{ik}^{ik} + \bar{h}_{jk}^{jk}$$
$$+ \bar{h}_{ab}^{ab} + \bar{h}_{ac}^{ac} + \bar{h}_{bc}^{bc}$$
$$- \bar{h}_{ai}^{ai} - \bar{h}_{aj}^{aj} - \bar{h}_{ak}^{ak}$$
$$- \bar{h}_{bi}^{bi} - \bar{h}_{bj}^{bj} - \bar{h}_{bk}^{bk} \tag{4.12}$$
$$- \bar{h}_{ci}^{ci} - \bar{h}_{cj}^{cj} - \bar{h}_{ck}^{ck},$$

and

$$\langle \Phi_{ijk}^{abc} | \bar{H}_3^{(\mathrm{SD})} | \Phi_{ijk}^{abc} \rangle = - \bar{h}_{aij}^{iaj} - \bar{h}_{aik}^{iak} - \bar{h}_{ajk}^{jak}$$
$$- \bar{h}_{bij}^{ibj} - \bar{h}_{bik}^{ibk} - \bar{h}_{bjk}^{jbk}$$
$$- \bar{h}_{cij}^{icj} - \bar{h}_{cik}^{ick} - \bar{h}_{cjk}^{jck}$$
$$+ \bar{h}_{aib}^{iab} + \bar{h}_{aic}^{iac} + \bar{h}_{bic}^{ibc} \tag{4.13}$$
$$+ \bar{h}_{ajb}^{jab} + \bar{h}_{ajc}^{jac} + \bar{h}_{bjc}^{jbc}$$
$$+ \bar{h}_{akb}^{kab} + \bar{h}_{akc}^{kac} + \bar{h}_{bkc}^{kbc},$$

where the little $\bar{h}$ values denote the antisymmetrized matrix elements of the one-, two-, and three-particle operators of the similarity transformed Hamiltonian, such that

$$\bar{H}_1^{(\mathrm{SD})} = \bar{h}_p^q \, N[a^p a_q], \tag{4.14}$$

$$\bar{H}_2^{(\mathrm{SD})} = \tfrac{1}{4} \bar{h}_{pq}^{rs} \, N[a^p a^q a_s a_r], \tag{4.15}$$

and

$$\bar{H}_3^{(\mathrm{SD})} = \tfrac{1}{36} \bar{h}_{pqr}^{stu} \, N[a^p a^q a^r a_u a_t a_s]. \tag{4.16}$$

The diagrams and corresponding algebraic expressions defining these operators, where little $\bar{h}$ amplitudes are expanded into their corresponding spin-integrated internal contractions among operators F_N, V_N, T_1, and T_2, are shown in Tables .3 through .7 in the Appendix.

Even though expressions for the remaining off-diagonal matrix elements of $\bar{H}_{N,\text{open}}^{(\text{SD})}$ in the N-electron basis can be explicitly written and implemented in software, we will take a slightly different route in order to simplify the computational implementation. But before we explain this alternative way, let us first review the well-known Slater rules (cf. [265] for a thorough introduction into the fundamentals of modern QC).

The expressions defining the non-zero matrix elements of any operator containing up to two-body terms are given by 5 cases, which correspond to the diagonal and off-diagonal elements of their one- and two-body components, and which are characterized by determinants with up to two different spin-orbitals among each other. For example, if we consider the Slater rules on the bare Hamiltonian, H, the matrix elements of its one-body component, Z, are given by

$$\langle \Phi_k | Z | \Phi_k \rangle = \sum_p \langle p|z|p \rangle = z_p^p, \tag{4.17}$$

and

$$\langle \Phi_k | Z\, a^a a_i | \Phi_k \rangle = (-1)^\rho \langle a|z|i \rangle = (-1)^\rho z_a^i, \tag{4.18}$$

where $(-1)^\rho$ is the sign of the permutation parity, ρ, resulting from the action of the creation (annihilation) operators a^a (a_i) on the determinant $|\Phi_k\rangle$. The first and second expressions correspond to the diagonal (i.e., the case where the bra and ket states are the same) and off-diagonal terms (i.e. the case where the bra and ket states differ by 1 spin orbital), respectively. Similarly, the matrix elements of the two-body components, V, of H, are defined by the cases

$$\langle \Phi_k | V | \Phi_k \rangle = \sum_{p,q} \langle pq|v|pq \rangle = \tfrac{1}{2} v_{pq}^{pq}, \tag{4.19}$$

$$\langle\Phi_k|V\,a^a a_i|\Phi_k\rangle = (-1)^\rho \sum_p \langle pa|v|pi\rangle = v_{pa}^{pi}, \tag{4.20}$$

and

$$\langle\Phi_k|V\,a^a a_i a^b a_j|\Phi_k\rangle = (-1)^\rho \langle ab|v|ij\rangle = v_{ab}^{ij}, \tag{4.21}$$

which correspond to the diagonal, and two off-diagonal terms, where one and two different spin orbitals distinguish the bra and ket states. In all cases, the p and q indices represent the occupied orbitals in the determinant $|\Phi_k\rangle$, belonging to the space of single-particle functions, $\mathcal{V}_1$, such that

$$|\Phi_k\rangle = |\{p_1\ldots p_N\}\rangle = |\{q_1\ldots q_N\}\rangle = \prod_p a^p |0\rangle = \prod_q a^q |0\rangle, \tag{4.22}$$

where $|0\rangle$ is the true vacuum.

An analogous formulation can be derived if one shifts the creation and annihilation operators, a^p or a_q, toward the Fermi vacuum, $|\Phi\rangle$, such that a new set of creation and annihilation operators, $\bar{a}^p$ or $\bar{a}_q$, act on a reference determinant, $|\Phi\rangle$, defining the Fermi vacuum, in the same way the operators a^p and a_q act on true vacuum, $|0\rangle$. There is a correspondence between true and Fermi vacuum creation and annihilation operators given by

$$\bar{a}^a = a^a, \qquad\qquad\qquad \bar{a}_a = a_a, \tag{4.23}$$

$$\bar{a}^i = a_i, \qquad\qquad\qquad \bar{a}_i = a^i, \tag{4.24}$$

where the indices i and a represent occupied and unoccupied orbitals in the Fermi vacuum, $|\Phi\rangle$, respectively. Using these mappings one can rewrite any Slater determinant, such that

$$|\Phi_k\rangle = \bar{a}^{a_1}\bar{a}^{i_1}\ldots\bar{a}^{a_n}\bar{a}^{i_n}|\Phi\rangle, \tag{4.25}$$

which allows us to redefine the original Slater rules such that instead of sums over occupied indices in $|\Phi_k\rangle$, we sum over indices defining the excitation with respect to $|\Phi\rangle$. If we apply this scheme in the computation of matrix elements of $\bar{H}$,[1] we obtain the

---

[1] From this point onward, we will drop the notation $\bar{H}_N^{(SD)}$ in favor of $\bar{H}$ for the sake of conciseness.

following expressions:

$$\langle\Phi_k|\bar{H}_1|\Phi_k\rangle = \langle\Phi|\,\bar{a}_k\bar{a}_c\bar{a}_j\bar{a}_b\bar{a}_i\bar{a}_a\,\bar{h}_p^q\bar{a}^p\bar{a}_q\,\bar{a}^a\bar{a}^i\bar{a}^b\bar{a}^j\bar{a}^c\bar{a}^k\,|\Phi\rangle$$

$$= \sum_{p\in\{i,j,k,a,b,c\}} \bar{h}_p^p, \tag{4.26}$$

and

$$\langle\Phi_k|\bar{H}_1\bar{a}^{p'}\bar{a}_{q'}|\Phi_k\rangle = (-1)^\rho\bar{h}_{p'}^{q'}, \tag{4.27}$$

for the one-body component of $\bar{H}$,

$$\langle\Phi_k|\bar{H}_2|\Phi_k\rangle = \sum_{p,q\in\{i,j,k,a,b,c\}} \bar{h}_{pq}^{pq}, \tag{4.28}$$

$$\langle\Phi_k|\bar{H}_2\bar{a}^{q'}\bar{a}_{s'}|\Phi_k\rangle = (-1)^\rho\sum_{p\in\{i,j,k,a,b,c\}} \bar{h}_{pq'}^{ps'}, \tag{4.29}$$

and

$$\langle\Phi_k|\bar{H}_2\bar{a}^{p'}\bar{a}_{r'}\bar{a}^{q'}\bar{a}_{s'}|\Phi_k\rangle = (-1)^\rho\bar{h}_{pq}^{rs} \tag{4.30}$$

for the two-body component of $\bar{H}$,

$$\langle\Phi_k|\bar{H}_3|\Phi_k\rangle = \sum_{p,q,r\in\{i,j,k,a,b,c\}} \bar{h}_{pqr}^{pqr}, \tag{4.31}$$

$$\langle\Phi_k|\bar{H}_3\bar{a}^{r'}\bar{a}_{u'}|\Phi_k\rangle = (-1)^\rho\sum_{p,q\in\{i,j,k,a,b,c\}} \bar{h}_{pqr'}^{pqu'}, \tag{4.32}$$

$$\langle\Phi_k|\bar{H}_3\bar{a}^{q'}\bar{a}_{t'}\bar{a}^{r'}\bar{a}_{u'}|\Phi_k\rangle = (-1)^\rho\sum_{p\in\{i,j,k,a,b,c\}} \bar{h}_{pq'r'}^{pt'u'}, \tag{4.33}$$

and

$$\langle\Phi_k|\bar{H}_3\bar{a}^{p'}\bar{a}_{s'}\bar{a}^{q'}\bar{a}_{t'}\bar{a}^{r'}\bar{a}_{u'}|\Phi_k\rangle = (-1)^\rho\bar{h}_{p'q'r'}^{s't'u'}, \tag{4.34}$$

for the three-body component of $\bar{H}$.

In these expressions, however, we must pay particular attention to the little $\bar{h}$ amplitudes, defining the various components of the similarity transformed Hamiltonian, which have to be pre-computed using the expressions shown in Tables .3 through .7 in the Appendix. If we take the computer resource requirements associated with keeping

these amplitudes in memory into consideration, we quickly realize that higher-order components of $\bar{H}$ are impractical in our new approach. For example, the matrix representation of the three-body component of $\bar{H}$ have memory requirements that scale as n^6, where $n = n_o + n_u$ is the number of orbitals defining the single-particle basis and n_o (n_u) is the number of occupied (unoccupied) orbitals. Thus, to generate the three-body components entering the matrix elements of $\bar{H}$, we decided, instead, to calculate the corresponding little $\bar{h}_{pqr}^{stu}$ values on the fly. To do so, one must catch all cases resulting from the computation of the diagonal and off-diagonal matrix elements of $\bar{H}_3$, which correspond to zero, one, two, and three spin-orbital differences between the bra and ket Slater determinants. Since we are interested in matrix elements were both the bra and ket states are triply excited determinants, the only terms contributing to these matrix elements result from $V_N T_2$ and are detailed in Table .7 in the Appendix.

In order to calculate these terms, one must deduct the indices corresponding to the external lines of $V_N T_2$ diagrams and contract over the single line connecting both terms. The sign of the element is given by the parity of the permutation that maximizes the overlap between the bra and ket states, as well as the the nature of the internal contracting line (i.e., whether the internal contraction between V_N and T_2 runs over a particle, which leads to a positive sign, or hole, which leads to a negative sign). Also, since we are computing these matrix elements based on the information of stochasticly generated lists of determinants, where indices representing occupied spin-orbitals are organized in ascending order, one has to be careful when creating the corresponding excitations, in order to keep the phase factor coherent. In the spin-integrated CC codes that we use in our group, we arrange excitations such that α and β orbitals are ordered by increasing number among each of the spin cases, and by letting the α orbitals occupy the first positions while the β orbitals occupy the last positions. For example, a triply excited determinant where two α and one β orbitals are excited from occupied orbitals to unoccupied orbitals is defined as $|\Phi_{ij\tilde{k}}^{ab\tilde{c}}\rangle$, where $i < j$ and $a < b$ and $\tilde{k}$ and $\tilde{c}$ range the

space of occupied and unoccupied orbitals, respectively. On the other hand, an excited determinant given by $|\Phi_{i\tilde{j}\tilde{k}}^{a\tilde{b}\tilde{c}}\rangle$, is organized such that $\tilde{j} < \tilde{k}$ and $\tilde{b} < \tilde{c}$, and i and a range over the occupied and unoccupied orbitals, respectively.

The algorithm to compute the matrix elements of $\langle\Phi_{k'}|(V_N T_2)_C|\Phi_k\rangle$, where $\langle\Phi_{k'}|$ and $|\Phi_k\rangle$ are triply excited determinants captured by i-CIQMC is illustrated as decision trees in Figures 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7, and is summarized as follows:

1. First, find the number of different spin-orbitals between $\langle\Phi_{k'}|$ and $|\Phi_k\rangle$. If more than zero, compute the parity of the permutation required to ensure maximum overlap. If larger than 3, return with zero.

2. If both $\langle\Phi_{k'}|$ and $|\Phi_k\rangle$ are the same, i.e., if calculating diagonal elements where $|\Phi_k\rangle = \langle\Phi_{k'}|^{\dagger}$, find the spin case of $|\Phi_k\rangle$ (i.e., $\alpha/\alpha/\alpha$, $\alpha/\alpha/\beta$, $\beta/\beta/\beta$, or $\beta/\beta/\beta$) and use the corresponding expression in Section. A.2.1.

3. If $\langle\Phi_{k'}|$ and $|\Phi_k\rangle$ differ by one spin orbital, check whether the difference is over occupied or unoccupied orbitals. For each case, check if the difference is of α or β spin. Finally, for each of the four cases resulting from the previous considerations, find the spin case of $\langle\Phi_{k'}|$ and use the appropriate formula from Section. A.2.2.

4. If $\langle\Phi_{k'}|$ and $|\Phi_k\rangle$ differ by two spin orbitals, check whether the difference is over two occupied orbitals, one occupied and one occupied orbital, or two unoccupied orbitals. For each case of the occupied/occupied or unoccupied/unoccupied kind, check if the difference is of the α/α, α/β, or β/β spin. If the case is of the occupied/unoccupied kind, check whether the difference is of the α/α, α/β, β/α, or β/β kind. Finally, for each of the ten cases resulting from the previous considerations, find the spin case of $\langle\Phi_{k'}|$ and use the appropriate formula from Section. A.2.3.

5. If $\langle\Phi_{k'}|$ and $|\Phi_k\rangle$ differ by three spin orbitals, check whether the difference is over two occupied orbitals and one unoccupied, or one occupied and two unoccupied

orbitals. For each case, find the spin case of $\langle \Phi_{k'}|$. For each of the eight cases resulting from the previous considerations, find the spin case of $\langle \Phi_{k'}|$ and use the appropriate formula from Section. A.2.4.

The above algorithm is repeated for every matrix element $\langle \Phi_{k'}|(V_N T_2)_C|\Phi_k\rangle$ resulting from two nested loops ranging over the triply excited determinants, $|\Phi_{k'}\rangle, |\Phi_k\rangle \in \mathscr{H}^{(P)}$, obtained from i-FCIQMC simulations. Within this double loop, we proceed to multiply the matrix elements, individually, with the corresponding t_k, belonging to the triply excited component of $T^{(P)}$, and update the array holding the projections onto triply excited determinants, $\langle \Phi_{k'}|$, spanning the stochastically determined P space.

At this point, we shall note that this is a very early implementation of this procedure, and, thus, can be greatly improved in the future. For instance, instead of directly looping over pairs of determinants $\langle \Phi_{k'}|$ and $|\Phi_k\rangle$, and considering all the decision cases described above, one could envision an initial step, before the iterative stage of the CC(P) begins, where the lists of stochastically defined determinants, of which $\langle \Phi_{k'}|$ and $|\Phi_k\rangle$ are elements, are sorted according to the aforementioned spin and excitation cases required for computing matrix elements of $\bar{H}_3$. Doing so would allow the reduction of the if statements required inside the loops, which represent a considerable computational speedup. Also, one could take advantage of the parallelization of the resulting loops, using a shared, or distributed memory, framework such as OpenMP or Open MPI, which would improve the performance and scalability and these computations.

4.2 Optimizing the CAD-FCIQMC Approach by Exploiting On-The-Fly Contractions of the T_4 Operator

The key step in the CAD-FCIQMC method is the refactorization of the FCI wave function, where we use its CI coefficients, C_n, to construct cluster components, T_n, defining the CC wave function ansatz. This process, known as the cluster analysis of the FCI wave
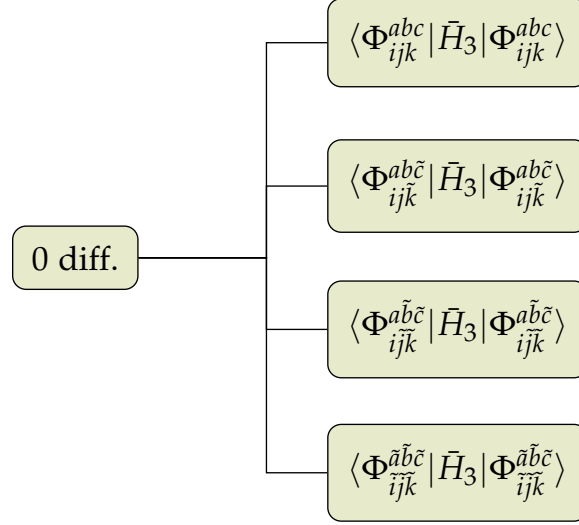
Figure 4.1: Decision tree showing the possible combinations of the spin-integrated bra and ket states resulting in non-zero diagonal matrix elements of $\bar{H}_3$.

function, is based on the original mapping between CI coefficients and cluster operators, introduced by Čížek in his seminal publication [25], and which is given by

$$
\begin{aligned}
C_1 &= T_1 \\
C_2 &= T_2 + \tfrac{1}{2}T_1^2 \\
C_3 &= T_3 + T_1 T_2 + \tfrac{1}{6}T_1^3 \\
C_4 &= T_4 + T_1 T_3 + \tfrac{1}{2}T_2^2 + \tfrac{1}{2}T_1^2 T_2 + \tfrac{1}{24}T_1^4, \text{ etc.}
\end{aligned}
\tag{4.35}
$$

Since we are interested in recovering the operators T_n out of the FCIQMC wave function, we need to invert these expressions such that instead of the exponential mapping of Eq. (4.35), where $\mathbb{1} + C_1 + C_2 + \cdots + C_N = e^{T_1 + T_2 + \cdots + T_N}$, we express the cluster operator in terms of the CI coefficients. The resulting logarithmic form, $\ln(\mathbb{1} + C_1 + C_2 + \cdots + C_N) = T_1 + T_2 + \cdots + T_N$, yields the following equivalences,

$$
\begin{aligned}
T_1 &= C_1 \\
T_2 &= C_2 - \tfrac{1}{2}C_1^2 \\
T_3 &= C_3 - C_1 C_2 + \tfrac{1}{3}C_1^3 \\
T_4 &= C_4 - C_1 C_3 + C_1^2 C_2 - \tfrac{1}{2}C_2^2 - \tfrac{1}{4}C_1^4, \text{ etc.}.
\end{aligned}
\tag{4.36}
$$
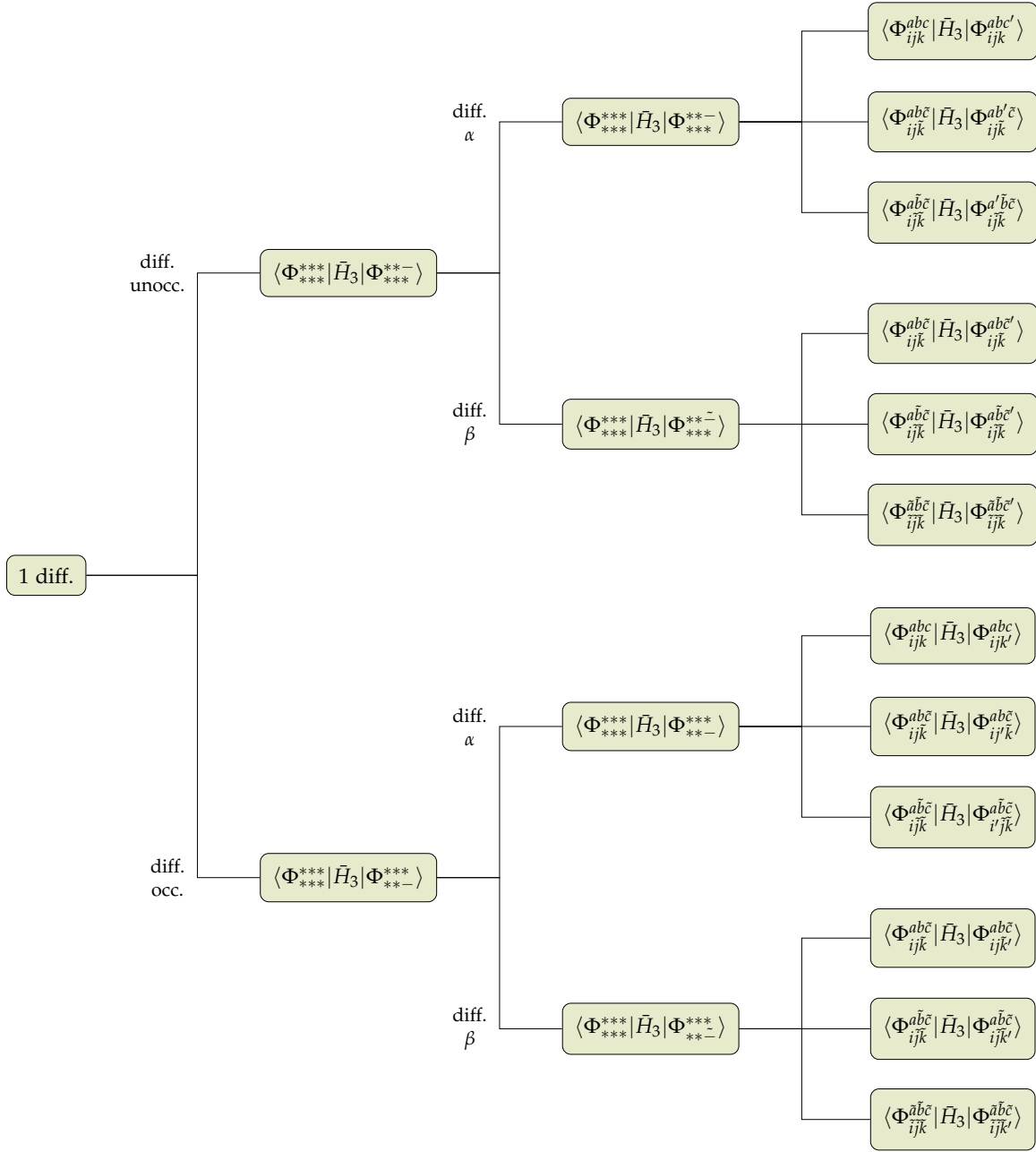
Figure 4.2: Decision tree showing the possible combinations of the spin-integrated bra and ket states resulting in non-zero off-diagonal matrix elements of $\bar{H}_3$ where one spin-orbital differs.
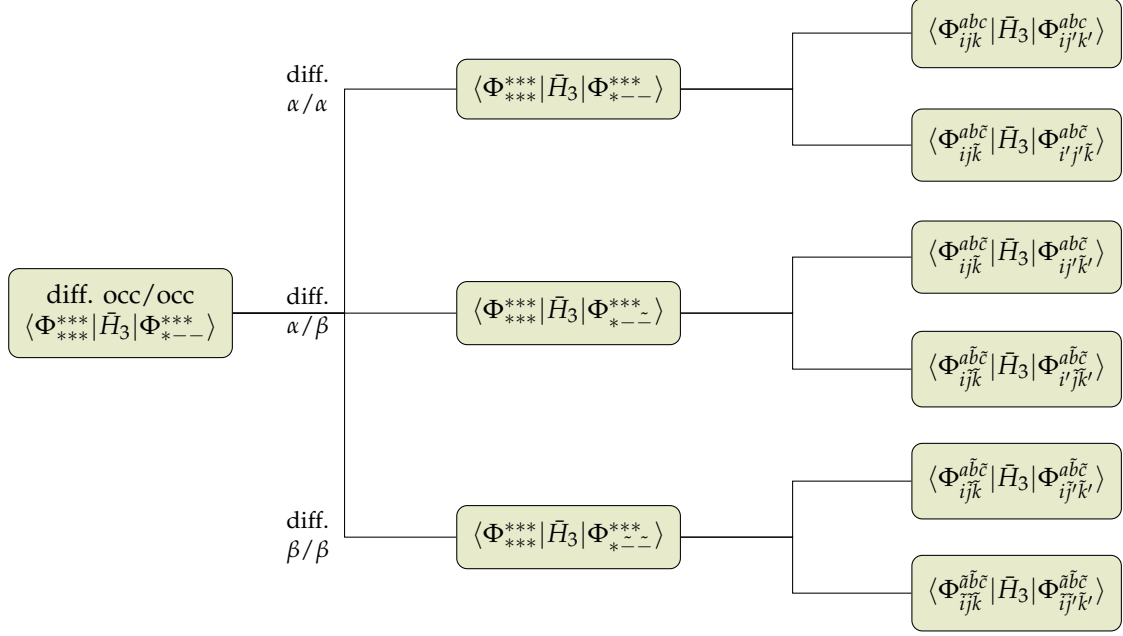
Figure 4.3: Decision tree showing the possible combinations of the spin-integrated bra and ket states resulting in non-zero off-diagonal matrix elements of $\bar{H}_3$ where two occupied spin-orbitals differ.

The Brandow diagrams and corresponding algebraic expressions, resulting from the above expressions for the cluster operators T_1, T_2, and T_3, are given in spin-integrated form in Tables 4.3, 4.4, and 4.5.

As we discussed in Section 2.6, the CAD-FCIQMC method exploits the fact that CC ground-state energies depend only on singly and doubly excited cluster components, that themselves only couple with their triply and quadruply excited counterparts. This means that, even though the FCIQMC simulations provide CI coefficients corresponding to excitations spanning the FCI space, we only need to pay attention to those CI coefficients required to build T_1, T_2, T_3, and T_4, namely, C_1, C_2, C_3, and C_4.

Even though, comparatively speaking, the T_3 and T_4 cluster components are much more manageable in terms of memory requirements than the higher-order T_n components with $n > 4$, they can still become memory bound since their associated memory scaling of $n_o^3 n_u^3$ and $n_o^4 n_u^4$, respectively, can quickly reach the multiple terabyte range when double precision float values are considered. Moreover, the CPU time step asso-
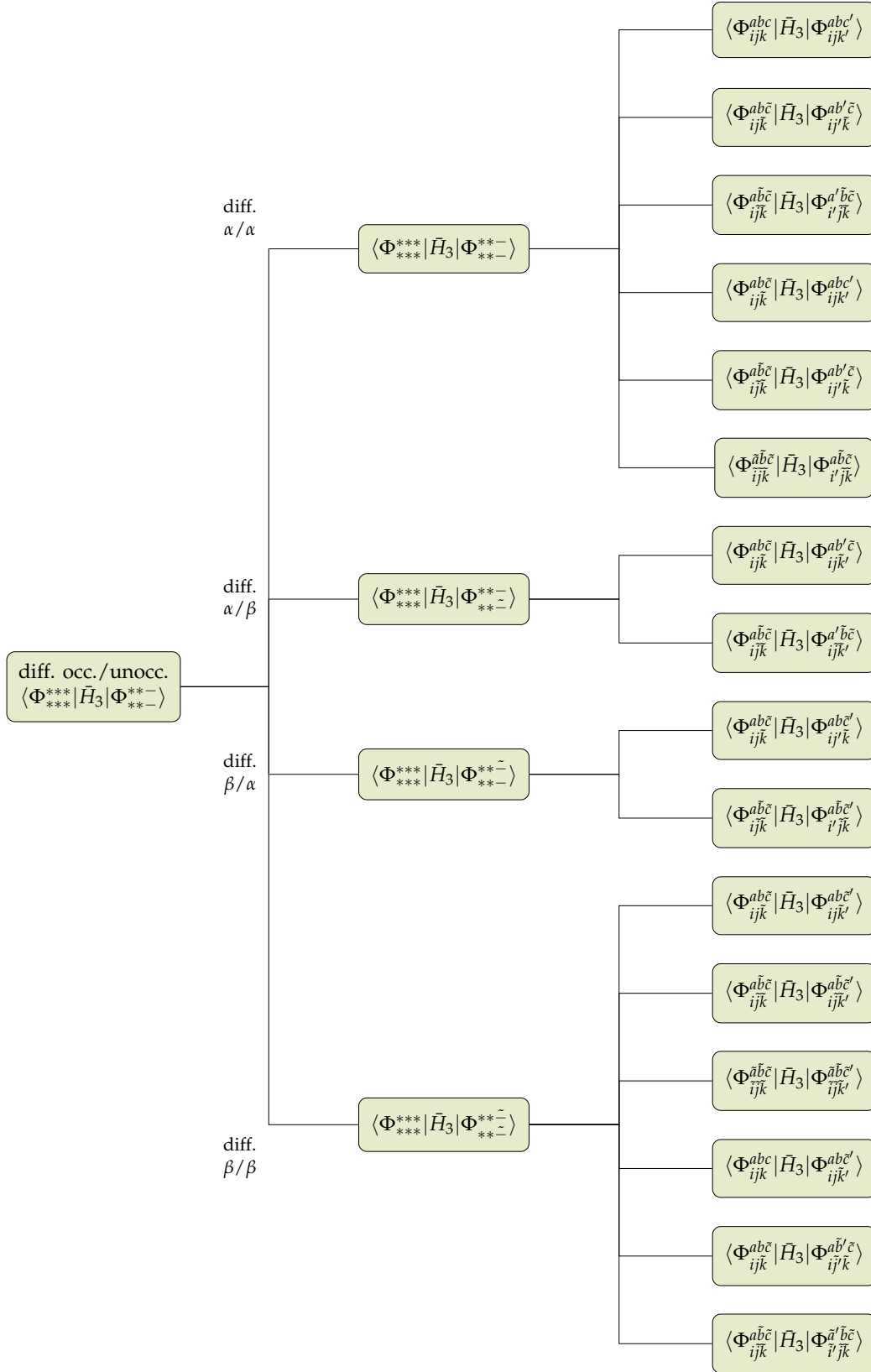
Figure 4.4: Decision tree showing the possible combinations of the spin-integrated bra and ket states resulting in non-zero off-diagonal matrix elements of $\bar{H}_3$ where one occupied and one unoccupied spin-orbitals differ.

107
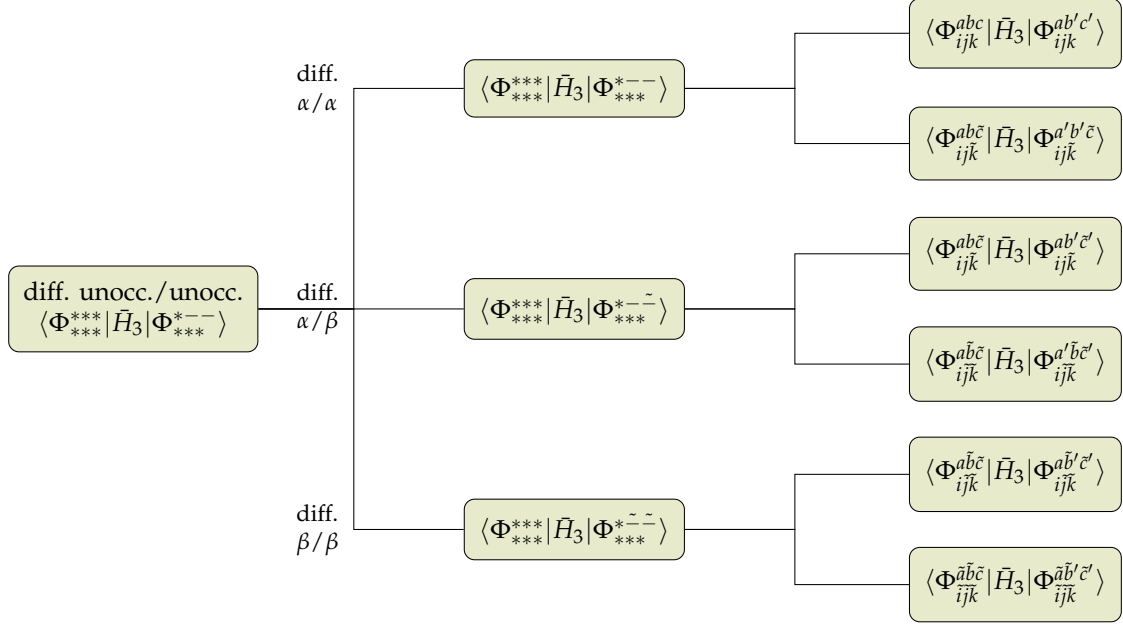
Figure 4.5: Decision tree showing the possible combinations of the spin-integrated bra and ket states resulting in non-zero off-diagonal matrix elements of $\bar{H}_3$ where two unoccupied spin-orbitals differ.

Table 4.3: Brandow diagrams and algebraic expressions corresponding to the singly excited cluster component equivalence, $T_1 = C_1$, for both spin cases of T_1.

$$t_a^i = c_a^i \qquad t_{\tilde{a}}^{\tilde{i}} = c_{\tilde{a}}^{\tilde{b}}$$

ciated with the cluster analysis of such large operators, which follow the same scaling laws, become too costly once more complex molecular systems are considered. To tackle these issues, our approach has been to take advantage of the sparsity of the FCIQMC wave function even further. Instead of generating the operators T_3 and T_4 entering Eqs. (2.89) and (2.90), we compute their corresponding projections, directly, by only considering those triply and quadruply excited determinants captured during i-CIQMC simulations. To explain this further, we will first write the term holding the largest
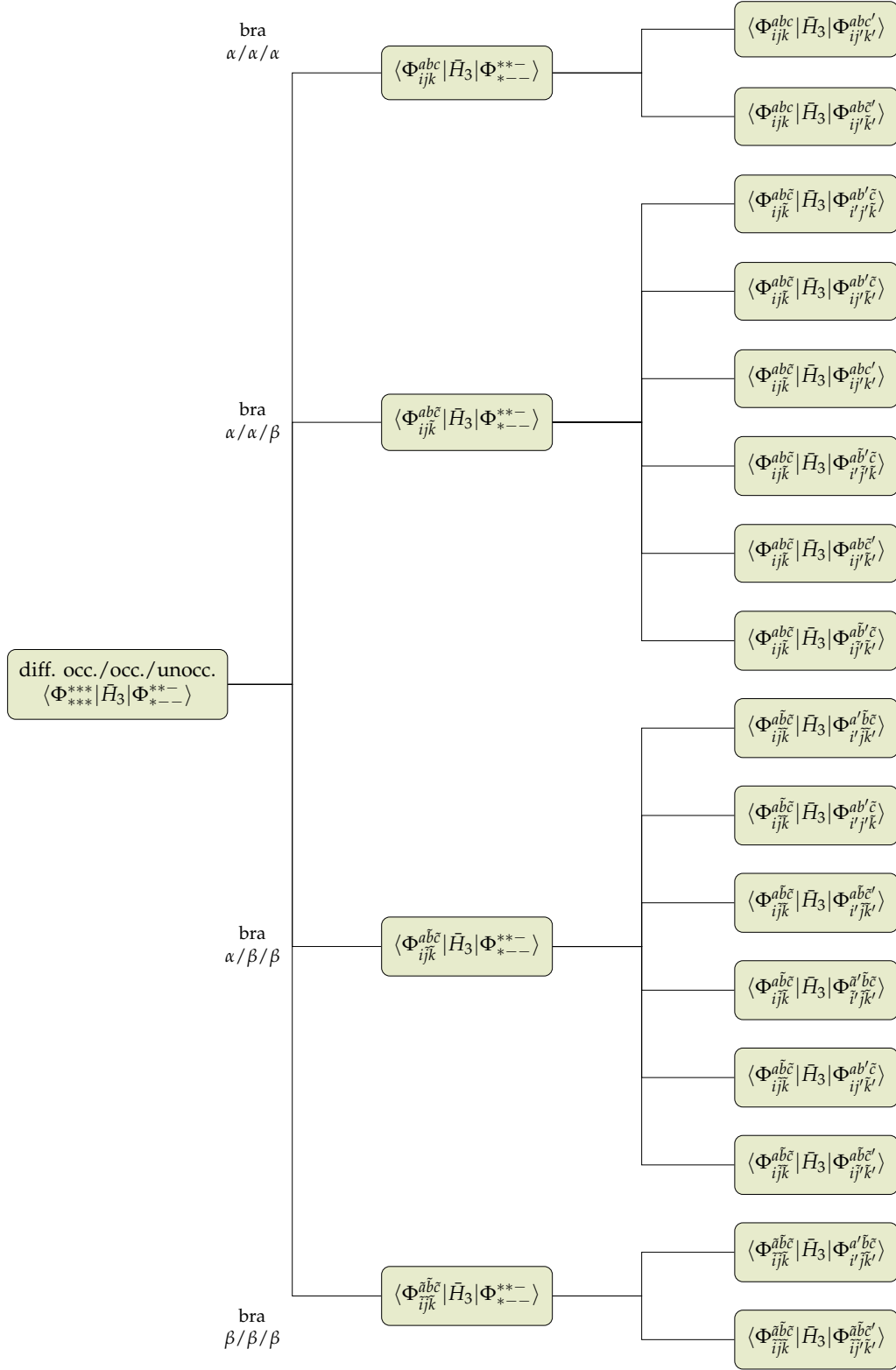
Figure 4.6: Decision tree showing the possible combinations of the spin-integrated bra and ket states resulting in non-zero off-diagonal matrix elements of $\bar{H}_3$ where two occupied and one unoccupied spin-orbitals differ.
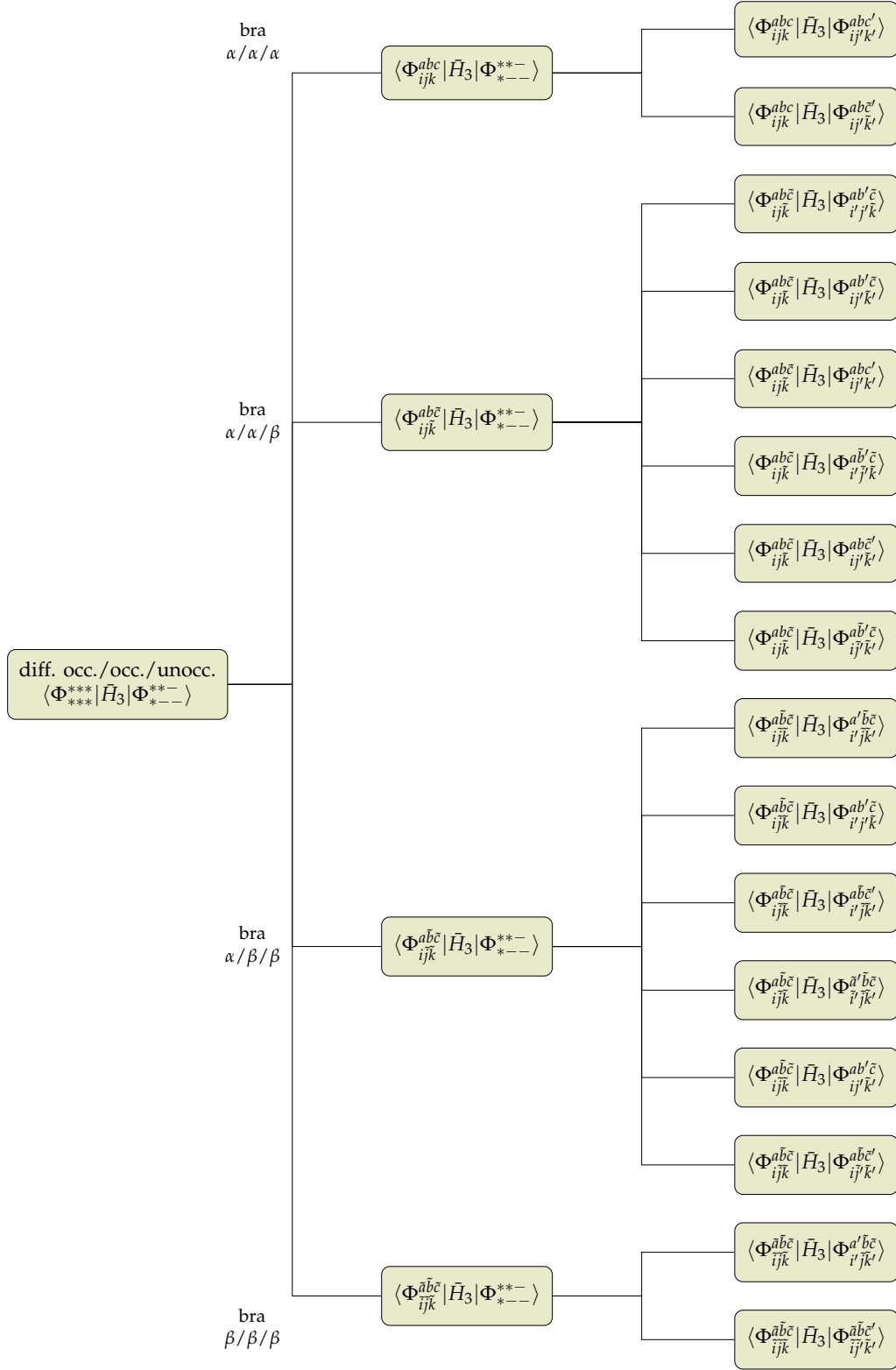
Figure 4.7: Decision tree showing the possible combinations of the spin-integrated bra and ket states resulting in non-zero off-diagonal matrix elements of $\bar{H}_3$ where one occupied and two unoccupied spin-orbitals differ.

Table 4.4: Brandow diagrams and algebraic expressions corresponding to the doubly excited cluster component equivalence, $T_2 = C_2 - \frac{1}{2}C_1^2$, for all three spin cases of T_2.

$$t^{ij}_{ab} = c^{ij}_{ab} - \mathscr{A}^{ij}c^i_a c^j_b \quad t^{i\tilde{j}}_{a\tilde{b}} = c^{i\tilde{j}}_{a\tilde{b}} - c^i_a c^{\tilde{j}}_{\tilde{b}} \quad t^{\tilde{i}\tilde{j}}_{\tilde{a}\tilde{b}} = c^{\tilde{i}\tilde{j}}_{\tilde{a}\tilde{b}} - \mathscr{A}^{\tilde{i}\tilde{j}}c^{\tilde{i}}_{\tilde{a}} c^{\tilde{j}}_{\tilde{b}}$$

cluster component in our considerations, namely

$$\langle \Phi^{ab}_{ij} | (V_N T_4)_C | \Phi \rangle = \langle \Phi^{ab}_{ij} | [V_N, T_4] | \Phi \rangle , \qquad (4.37)$$

which is the only term where T_4 is present in CAD-FCIQMC equations. By utilizing Eq. (4.35), we rewrite the quadruply excited cluster component, such that,

$$T_4 = C_4 - (T_1 T_3 + \tfrac{1}{2}T_2^2 + \tfrac{1}{2}T_1^2 T_2 + \tfrac{1}{24}T_1^4). \qquad (4.38)$$

Combining the above expressions, we can rearrange the commutator defining the projection of $(V_N T_4)_C |\Phi\rangle$ onto doubly excited determinants, such that

$$
\begin{aligned}
\langle \Phi^{ab}_{ij} | & [V_N, T_4] | \Phi \rangle \\
&= \langle \Phi^{ab}_{ij} | [V_N, C_4 - (T_1 T_3 + \tfrac{1}{2}T_2^2 + \tfrac{1}{2}T_1^2 T_2 + \tfrac{1}{24}T_1^4)] | \Phi \rangle \\
&= \langle \Phi^{ab}_{ij} | [V_N, C_4] | \Phi \rangle - \langle \Phi^{ab}_{ij} | [V_N, T_1 T_3] | \Phi \rangle \\
&\quad - \tfrac{1}{2} \langle \Phi^{ab}_{ij} | [V_N, T_2^2] | \Phi \rangle - \tfrac{1}{2} \langle \Phi^{ab}_{ij} | [V_N, T_1^2 T_2] | \Phi \rangle \\
&\quad - \tfrac{1}{24} \langle \Phi^{ab}_{ij} | [V_N, T_1^4)] | \Phi \rangle .
\end{aligned}
\qquad (4.39)
$$

Table 4.5: Brandow diagrams and algebraic expressions corresponding to the triply excited cluster component equivalence, $T_3 = C_3 - C_1 C_2 + \frac{1}{3} C_1^3$, for all four spin cases of T_3.

$$t^{ijk}_{abc} = c^{ijk}_{abc}$$
$$- \mathscr{A}^{k/ij}_{c/ab} c^{ij}_{ab} c^{k}_{c}$$
$$+ 2 \mathscr{A}^{ijk} c^{i}_{a} c^{j}_{b} c^{k}_{c}$$

$$t^{ij\tilde{k}}_{ab\tilde{c}} = c^{ij\tilde{k}}_{ab\tilde{c}}$$
$$- c^{ij}_{ab} c^{\tilde{k}}_{\tilde{c}}$$
$$- \mathscr{A}^{ij}_{ab} c^{i\tilde{k}}_{a\tilde{c}} c^{j}_{b}$$
$$+ 2 \mathscr{A}^{ij} c^{i}_{a} c^{j}_{b} c^{\tilde{k}}_{\tilde{c}}$$

$$t^{i\tilde{j}\tilde{k}}_{a\tilde{b}\tilde{c}} = c^{i\tilde{j}\tilde{k}}_{a\tilde{b}\tilde{c}}$$
$$- c^{i}_{a} c^{\tilde{j}\tilde{k}}_{\tilde{b}\tilde{c}}$$
$$- \mathscr{A}^{\tilde{j}\tilde{k}}_{\tilde{b}\tilde{c}} c^{i\tilde{j}}_{a\tilde{b}} c^{\tilde{k}}_{\tilde{c}}$$
$$+ 2 \mathscr{A}^{\tilde{j}\tilde{k}} c^{i}_{a} c^{\tilde{j}}_{\tilde{b}} c^{\tilde{k}}_{\tilde{c}}$$

$$t^{\tilde{i}\tilde{j}\tilde{k}}_{\tilde{a}\tilde{b}\tilde{c}} = c^{\tilde{i}\tilde{j}\tilde{k}}_{\tilde{a}\tilde{b}\tilde{c}}$$
$$- \mathscr{A}^{\tilde{k}/\tilde{i}\tilde{j}}_{\tilde{c}/\tilde{a}\tilde{b}} c^{\tilde{i}\tilde{j}}_{\tilde{a}\tilde{b}} c^{\tilde{k}}_{\tilde{c}}$$
$$+ 2 \mathscr{A}^{\tilde{i}\tilde{j}\tilde{k}} c^{\tilde{i}}_{\tilde{a}} c^{\tilde{j}}_{\tilde{b}} c^{\tilde{k}}_{\tilde{c}}$$

Despite the similarity of these equations to the ones resulting from the standard SR CCSDTQ theory when projections onto doubly excited determinants are considered (cf. Ref. [38–41]), where powers of the cluster operator appear within nested commutators in the Baker–Campbell–Hausdorf (BCH) expansion, and, thus, are represented in connected diagrams only, the terms in Eq. (4.39) are defined by a *single* commutator, even when products of multiple cluster components are present. This means that the diagrams representing these terms are no longer strictly connected, but also admit their partially disconnected counterparts. In general, a single commutator, mandates that diagrams with at least one connection between operators in the left and right hand sides of the commutator remain, or in other words, that completely disconnected diagrams must be removed. Therefore, in the case of Eq. (4.39), we must collect diagrams where the V_N operator connects to at least one cluster component. The spin-integrated terms, organized by the relevant spin cases, and which result from these considerations when applied to $\langle \Phi_{ij}^{ab} | [V_N, T_1 T_3] | \Phi \rangle$, $\frac{1}{2} \langle \Phi_{ij}^{ab} | [V_N, T_2^2] | \Phi \rangle$, $\frac{1}{2} \langle \Phi_{ij}^{ab} | [V_N, T_1^2 T_2] | \Phi \rangle$, and $\frac{1}{24} \langle \Phi_{ij}^{ab} | [V_N, T_1^4] | \Phi \rangle$, are shown in Tables 4.6, 4.7, 4.8, and 4.9, respectively.

The remaining and most memory-expensive term, $\langle \Phi_{ij}^{ab} | [V_N, C_4] | \Phi \rangle$, yields only connected diagrams, since it is linear in C_4, and, thus, its corresponding algebraic expression is is given by $\frac{1}{4} v_{mn}^{ef} c_{abef}^{ijmn}$, which results from contracting all four lines from the operator V_N with the C_4 vector. However, instead of performing these contractions over indices m, n, e, f, as it is usually done using matrix-matrix multiplication routines, we capitalize on the fact that the C_4 coefficients arriving from *i*-CIQMC simulations are quite sparse when compared against the total amount of quadruples in FCI, which allow us to compute only the reduced set of contractions where CI coefficients are not zero. To do this, we employ Algorithm 1, which consists in the following steps: First, we run a loop over each of the C_4 coefficients, with non-zero walker populations, encountered during *i*-FCIQMC simulations. Within it, we run a second loop that identifies double de-excitations out of the quadruply excited, ket, determinants accompanying C_4, that

Table 4.6: Spin-integrated terms belonging to the projection of the commutator, $[V_N, T_1 T_3]|\Phi\rangle$, onto double excited determinants $\langle\Phi_{ij}^{ab}|$, $\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|$, and $\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|$ corresponding to the α/α, α/β, and β/β spin cases. The subscripts A, B, C, and D denote the various spin cases associated with spin-integrated operators. The A subscript refers to the case where all excitations are of α spin, up to the many-body excitation rank n of the corresponding operator. Subsequently, the B subscript denotes the spin case where one excitation is of β spin while the remaining ones are of the α type. The remaining indices C and D increase the number of β-type excitations by one and two, respectively, such that two-body operators of the C type and three-body operators of the D type consist of only β excitations. The single commutator in this expressions mandates that at least one connected term remains in the diagrammatic expressions. Also, the corresponding weights have been dropped for conciseness.

$\langle\Phi_{ij}^{ab}| [V_N, T_1 T_3] |\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}| [V_N, T_1 T_3] |\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}| [V_N, T_1 T_3] |\Phi\rangle$
$\langle\Phi_{ij}^{ab}| (V_{N,A} T_{1A} T_{3A})_C |\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}| (V_{N,A} T_{1A} T_{3B})_C |\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}| (V_{N,A} T_{1A} T_{3C})_C |\Phi\rangle$
$\langle\Phi_{ij}^{ab}| (V_{N,B} T_{1B} T_{3A})_C |\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}| (V_{N,B} T_{1B} T_{3B})_C |\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}| (V_{N,B} T_{1A} T_{3D})_C |\Phi\rangle$
$\langle\Phi_{ij}^{ab}| (V_{N,B} T_{1A} T_{3B})_C |\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}| (V_{N,C} T_{1B} T_{3C})_C |\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}| (V_{N,B} T_{1B} T_{3C})_C |\Phi\rangle$
$\langle\Phi_{ij}^{ab}| (V_{N,C} T_{1B} T_{3B})_C |\Phi\rangle$	$\langle\Phi_{ij}^{ab}| (V_{N,C} T_{1B} T_{3C})_C |\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}| (V_{N,C} T_{1B} T_{3D})_C |\Phi\rangle$
$\langle\Phi_{ij}^{ab}| ((V_{N,A} T_{3A})_C T_{1A})_D |\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}| ((V_{N,A} T_{3A})_C T_{1B})_D |\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}| ((V_{N,A} T_{3B})_C T_{1B})_D |\Phi\rangle$
$\langle\Phi_{ij}^{ab}| ((V_{N,B} T_{3B})_C T_{1A})_D |\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}| ((V_{N,B} T_{3B})_C T_{1B})_D |\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}| ((V_{N,B} T_{3C})_C T_{1B})_D |\Phi\rangle$
$\langle\Phi_{ij}^{ab}| ((V_{N,C} T_{3C})_C T_{1A})_D |\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}| ((V_{N,C} T_{3C})_C T_{1B})_D |\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}| ((V_{N,C} T_{3D})_C T_{1B})_D |\Phi\rangle$
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}| ((V_{N,A} T_{3B})_C T_{1A})_D |\Phi\rangle$	
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}| ((V_{N,B} T_{3C})_C T_{1A})_D |\Phi\rangle$	
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}| ((V_{N,C} T_{3D})_C T_{1A})_D |\Phi\rangle$	

produce a doubly excited, bra, determinant (i.e., excitations such that $\langle\Phi_{ij}^{ab}|E_{ef}^{mn}|\Phi_{ijmn}^{abef}\rangle$), which we subsequently use to compute the matrix elements of V_N multiplying the corresponding CI coefficient. The result of running this algorithm, is an array of indices i, j, a, and b, with a much smaller memory scaling of $n_o^2 n_u^2$, and which contains the projection, $\langle\Phi_{ij}^{ab}|[V_N, C_4]|\Phi\rangle$. Moreover, since only a fraction of the quadruples are considered in this procedure, corresponding to those quadruply excited captured during *i*-CIQMC simulations, we save in the CPU time as well.

As we discussed above, the remaining terms in the projection shown in Eq. (4.39),

Table 4.7: Spin-integrated terms belonging to the projection of the commutator, $\frac{1}{2}[V_N, (T_2)^2]\,|\Phi\rangle$, onto double excited determinants $\langle\Phi_{ij}^{ab}|$, $\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|$, and $\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|$ corresponding to the α/α, α/β, and β/β spin cases. The subscripts A, B, C, and D denote the various spin cases associated with spin-integrated operators. The A subscript refers to the case where all excitations are of α spin, up to the many-body excitation rank n of the corresponding operator. Subsequently, the B subscript denotes the spin case where one excitation is of β spin while the remaining ones are of the α type. The remaining indices C and D increase the number of β-type excitations by one and two, respectively, such that two-body operators of the C type and three-body operators of the D type consist of only β excitations. The single commutator in this expressions mandates that at least one connected term remains in the diagrammatic expressions. Also, the corresponding weights have been dropped for conciseness.

$\langle\Phi_{ij}^{ab}|\,[V_N,(T_2)^2]\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,[V_N,(T_2)^2]\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,[V_N,(T_2)^2]\,|\Phi\rangle$
$\langle\Phi_{ij}^{ab}|\,(V_{N,A}T_{2A}T_{2A})_C\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,(V_{N,A}T_{2A}T_{2B})_C\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,(V_{N,A}T_{2B}T_{2B})_C\,|\Phi\rangle$
$\langle\Phi_{ij}^{ab}|\,(V_{N,B}T_{2A}T_{2B})_C\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,(V_{N,B}T_{2A}T_{2C})_C\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,(V_{N,B}T_{2B}T_{2C})_C\,|\Phi\rangle$
$\langle\Phi_{ij}^{ab}|\,(V_{N,C}T_{2B}T_{2B})_C\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,(V_{N,B}T_{2B}T_{2B})_C\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,(V_{N,C}T_{2C}T_{2C})_C\,|\Phi\rangle$
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,(V_{N,C}T_{2B}T_{2C})_C\,|\Phi\rangle$	
$\langle\Phi_{ij}^{ab}|\,((V_{N,A}T_{2A})_C T_{2A})_D\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,((V_{N,A}T_{2A})_C T_{2B})_D\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,((V_{N,A}T_{2A})_C T_{2C})_D\,|\Phi\rangle$
$\langle\Phi_{ij}^{ab}|\,((V_{N,B}T_{2B})_C T_{2A})_D\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,((V_{N,B}T_{2B})_C T_{2B})_D\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,((V_{N,B}T_{2B})_C T_{2C})_D\,|\Phi\rangle$
$\langle\Phi_{ij}^{ab}|\,((V_{N,C}T_{2C})_C T_{2A})_D\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,((V_{N,C}T_{2C})_C T_{2B})_D\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,((V_{N,C}T_{2C})_C T_{2C})_D\,|\Phi\rangle$

Algorithm 1: Computation of $\langle\Phi_{ij}^{ab}|\,(V_N C_4)_C\,|\Phi\rangle$. The elements of the set Q consist of the pairs $(c, |\Phi_q\rangle)$ defining the list of quadruply excited components with non-zero walker populations in i-FCIQMC simulations.

ComputeVC4

    inputs: two-body integrals defining V_N; pairs of CI coefficients and
           quadruply excited determinants $(c, |\Phi_q\rangle) \in Q$

    output: doubly excited projection vector $P_2^* = p_{ij}^{*ab}$

    foreach $(c, |\Phi_q\rangle) \in Q$ **do**
        foreach $\langle\Phi_{ij}^{ab}|$ *such that* $\langle\Phi_{ij}^{ab}|V_N|\Phi_q\rangle \neq 0$ **do**
            $p_{ij}^{*ab} \leftarrow c\,\langle\Phi_{ij}^{ab}|V_N|\Phi_q\rangle$

    return P_2^*

Table 4.8: Spin-integrated terms belonging to the projection of the commutator, $\frac{1}{2}[V_N, (T_1)^2 T_2] |\Phi\rangle$, onto double excited determinants $\langle\Phi_{ij}^{ab}|$, $\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|$, and $\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|$ corresponding to the α/α, α/β, and β/β spin cases. The subscripts A, B, C, and D denote the various spin cases associated with spin-integrated operators. The A subscript refers to the case where all excitations are of α spin, up to the many-body excitation rank n of the corresponding operator. Subsequently, the B subscript denotes the spin case where one excitation is of β spin while the remaining ones are of the α type. The remaining indices C and D increase the number of β-type excitations by one and two, respectively, such that two-body operators of the C type and three-body operators of the D type consist of only β excitations. The single commutator in this expressions mandates that at least one connected term remains in the diagrammatic expressions. Also, the corresponding weights have been dropped for conciseness.

$\langle\Phi_{ij}^{ab}\| [V_N, (T_1)^2 T_2] \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| [V_N, (T_1)^2 T_2] \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| [V_N, (T_1)^2 T_2] \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| (V_{N,A} T_{1A} T_{1A} T_{2A})_C \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| (V_{N,A} T_{1A} T_{1A} T_{2B})_C \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| (V_{N,B} T_{1A} T_{1B} T_{2C})_C \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| (V_{N,B} T_{1A} T_{1A} T_{2B})_C \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| (V_{N,B} T_{1A} T_{1A} T_{2C})_C \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| (V_{N,B} T_{1B} T_{1B} T_{2B})_C \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| (V_{N,B} T_{1A} T_{1B} T_{2A})_C \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| (V_{N,B} T_{1A} T_{1B} T_{2B})_C \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| (V_{N,C} T_{1B} T_{1B} T_{2C})_C \|\Phi\rangle$
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| (V_{N,B} T_{1B} T_{1B} T_{2A})_C \|\Phi\rangle$	
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| (V_{N,C} T_{1B} T_{1B} T_{2B})_C \|\Phi\rangle$	
$\langle\Phi_{ij}^{ab}\| ((V_{N,A} T_{2A})_C T_{1A} T_{1A})_D \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,A} T_{2A})_C T_{1A} T_{1B})_D \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| ((V_{N,A} T_{2A})_C T_{1B} T_{1B})_D \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| ((V_{N,B} T_{2B})_C T_{1A} T_{1A})_D \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,B} T_{2B})_C T_{1A} T_{1B})_D \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| ((V_{N,B} T_{2B})_C T_{1B} T_{1B})_D \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| ((V_{N,C} T_{2C})_C T_{1A} T_{1A})_D \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,C} T_{2C})_C T_{1A} T_{1B})_D \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| ((V_{N,C} T_{2C})_C T_{1B} T_{1B})_D \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| ((V_{N,A} T_{1A} T_{1A})_C T_{2A})_D \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,A} T_{1A} T_{1A})_C T_{2B})_D \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| ((V_{N,A} T_{1A} T_{1A})_C T_{2C})_D \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| ((V_{N,B} T_{1A} T_{1B})_C T_{2A})_D \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,B} T_{1A} T_{1B})_C T_{2B})_D \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| ((V_{N,B} T_{1A} T_{1B})_C T_{2C})_D \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| ((V_{N,C} T_{1B} T_{1B})_C T_{2A})_D \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,C} T_{1B} T_{1B})_C T_{2B})_D \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| ((V_{N,C} T_{1B} T_{1B})_C T_{2C})_D \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| ((V_{N,A} T_{1A} T_{2A})_C T_{1A})_D \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,A} T_{1A} T_{2A})_C T_{1B})_D \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| ((V_{N,B} T_{1B} T_{2B})_C T_{1B})_D \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| ((V_{N,B} T_{1A} T_{2B})_C T_{1A})_D \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,B} T_{1A} T_{2B})_C T_{1B})_D \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| ((V_{N,C} T_{1B} T_{2C})_C T_{1B})_D \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| ((V_{N,B} T_{1B} T_{2A})_C T_{1A})_D \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,B} T_{1B} T_{2B})_C T_{1A})_D \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| ((V_{N,B} T_{1A} T_{2C})_C T_{1B})_D \|\Phi\rangle$
$\langle\Phi_{ij}^{ab}\| ((V_{N,C} T_{1B} T_{2B})_C T_{1A})_D \|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,C} T_{1B} T_{2C})_C T_{1A})_D \|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}\| ((V_{N,A} T_{1A} T_{2B})_C T_{1B})_D \|\Phi\rangle$
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,B} T_{1B} T_{2A})_C T_{1B})_D \|\Phi\rangle$	
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,C} T_{1B} T_{2B})_C T_{1B})_D \|\Phi\rangle$	
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,A} T_{1A} T_{2B})_C T_{1A})_D \|\Phi\rangle$	
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}\| ((V_{N,B} T_{1A} T_{2C})_C T_{1A})_D \|\Phi\rangle$	

Table 4.9: Spin-integrated terms belonging to the projection of the commutator, $\frac{1}{24}[V_N,(T_1)^4]\,|\Phi\rangle$, onto double excited determinants $\langle\Phi_{ij}^{ab}|$, $\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|$, and $\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|$ corresponding to the α/α, α/β, and β/β spin cases. The subscripts A, B, C, and D denote the various spin cases associated with spin-integrated operators. The A subscript refers to the case where all excitations are of α spin, up to the many-body excitation rank n of the corresponding operator. Subsequently, the B subscript denotes the spin case where one excitation is of β spin while the remaining ones are of the α type. The remaining indices C and D increase the number of β-type excitations by one and two, respectively, such that two-body operators of the C type and three-body operators of the D type consist of only β excitations. The single commutator in this expressions mandates that at least one connected term remains in the diagrammatic expressions. Also, the corresponding weights have been dropped for conciseness.

$\langle\Phi_{ij}^{ab}|\,[V_N,(T_1)^4]\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,[V_N,(T_1)^4]\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,[V_N,(T_1)^4]\,|\Phi\rangle$
$\langle\Phi_{ij}^{ab}|\,(V_{N,A}(T_{1A})^4)_C\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,(V_{N,B}(T_{1A})^2(T_{1B})^2)_C\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,(V_{N,C}(T_{1B})^4)_C\,|\Phi\rangle$
$\langle\Phi_{ij}^{ab}|\,((V_{N,A}(T_{1A})_C^2(T_{1A})^2)_D\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,((V_{N,A}(T_{1A})^2)_C T_{1A}T_{1B})_D\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,((V_{N,A}(T_{1A})^2)_C(T_{1B})^2)_D\,|\Phi\rangle$
$\langle\Phi_{ij}^{ab}|\,((V_{N,B}T_{1A}T_{1B})_C(T_{1A})^2)_D\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,((V_{N,B}T_{1A}T_{1B})_C T_{1A}T_{1B})_D\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,((V_{N,B}T_{1A}T_{1B})_C(T_{1B})^2)_D\,|\Phi\rangle$
$\langle\Phi_{ij}^{ab}|\,((V_{N,C}(T_{1B})^2)_C(T_{1A})^2)_D\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,((V_{N,C}(T_{1B})^2)_C T_{1A}T_{1B})_D\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,((V_{N,C}(T_{1B})^2)_C(T_{1B})^2)_D\,|\Phi\rangle$
$\langle\Phi_{ij}^{ab}|\,((V_{N,A}(T_{1A})^3)_C T_{1A})_D\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,((V_{N,A}(T_{1A})^3)_C T_{1B})_D\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,((V_{N,B}T_{1A}(T_{1B})^2)_C T_{1B})_D\,|\Phi\rangle$
$\langle\Phi_{ij}^{ab}|\,((V_{N,B}(T_{1A})^2T_{1B})_C T_{1A})_D\,|\Phi\rangle$	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,((V_{N,B}(T_{1A})^2T_{1B})_C T_{1B})_D\,|\Phi\rangle$	$\langle\Phi_{\tilde{i}\tilde{j}}^{\tilde{a}\tilde{b}}|\,((V_{N,C}(T_{1B})^3)_C T_{1B})_D\,|\Phi\rangle$
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,((V_{N,B}T_{1A}(T_{1B})^2)_C T_{1A})_D\,|\Phi\rangle$	
	$\langle\Phi_{i\tilde{j}}^{a\tilde{b}}|\,((V_{N,C}(T_{1B})^3)_C T_{1A})_D\,|\Phi\rangle$	

involve partially disconnected terms, and, thus, extra steps are needed. First, we need to separate the completely connected terms from those with partial disconnections. The result of this process is shown in Tables 4.6, 4.7, 4.8, and 4.9, where the first few rows of expressions correspond to the spin-integrated connected terms, and the second row involves the partially disconnected ones. The connected terms correspond to the diagrams resulting from regular SR CCSDTQ equations when projected onto doubly excited determinants. Therefore, we use the relevant CCSDTQ subroutines to calculate these terms, efficiently, without the need for further manipulation. On the other hand, the partially disconnected terms are not included in regular CCSDTQ consideration, and so, we will now focus on these new terms.

The Hugenholtz and Brandow diagrams, along with their corresponding algebraic

expressions, for the spin-integrated projection of the partially disconnected terms onto doubly excited determinants comprising only α excitations are shown in Tables .8, .9, .10, .11 in the Appendix. By examining these diagrams, one can quickly realize that all diagrams consist of a connected and disconnected pieces, and as such can be easily factorized. All connected elements in the diagrams correspond to either a vacuum or singly excited piece, while the disconnected parts are always composed of uncontracted cluster operators. Therefore, the factorized expressions have two general forms. The first one comprises the terms $\frac{1}{2}T_1^2 + T_2$ scaled by the vacuum pieces resulting from $\langle\Phi|(V_N\left[\frac{1}{2}T_1^2 + T_2\right])_C|\Phi\rangle$. The second form consists in the outer product of T_1 with the projection of the remaining connected terms onto singly excited determinants, given by $\langle\Phi_i^a|(V_N\left[\frac{1}{6}T_1^3 + T_1 T_2 + T_3\right])_C|\Phi\rangle$. The expressions resulting from these considerations and defining all the spin-integrated and factorized partially disconnected terms are given in Tables 4.10, 4.11, and 4.12.

In summary, to optimize the most expensive term given by the projection of $(V_N T_4)_C|\Phi\rangle$ onto doubly excited determinants, we have decomposed the corresponding commutator into the five terms given by Eq. (4.39). The one term containing quadruply excited CI coefficients was computed by considering only those quadruply excited determinants obtained from i-FCIQMC simulations, with their corresponding walker populations, using Algorithm 1. This approach results in considerable computational savings since, as discussed in Section 2.6, the number of quadruply excited determinants captured by i-FCIQMC is much smaller relative to the total amount. The remaining terms in Eq. (4.39), which contain the smaller T_1, T_2, and T_3 clusters, were computed by factorizing the connected and disconnected parts of the resulting projections onto doubly excited determinants. This allowed us to take advantage of the already available and highly optimized routines, which are part of standard SR CCSDT and CCSDTQ software implementations, for computing the connected parts of the equations shown in Tables 4.6, 4.7, 4.8, 4.9, 4.10, 4.11, and 4.12. Subsequently, the final expressions Tables

Table 4.10: Spin-integrated and factorized terms resulting from the semi-disconnected diagrams resulting from Eq. (4.39) projected onto doubly excited determinants of the α/α type.

	$(t_{ab}^{ij} + \mathscr{A}_{ij} t_a^i t_b^j)\times$	$\mathscr{A}_{ij}^{ab} t_a^i \times$
zero-body	$\langle\Phi|(V_{N,A}T_{2A})_C|\Phi\rangle$	–
	$\langle\Phi|(V_{N,B}T_{2B})_C|\Phi\rangle$	–
	$\langle\Phi|(V_{N,C}T_{2C})_C|\Phi\rangle$	–
	$\frac{1}{2}\langle\Phi|(V_{N,A}T_{1A}T_{1A})_C|\Phi\rangle$	–
	$\langle\Phi|(V_{N,B}T_{1A}T_{1B})_C|\Phi\rangle$	–
	$\frac{1}{2}\langle\Phi|(V_{N,C}T_{1B}T_{1B})_C|\Phi\rangle$	–
one-body	–	$\langle\Phi_j^b|(V_{N,A}T_{3A})_C|\Phi\rangle$
	–	$\langle\Phi_j^b|(V_{N,B}T_{3B})_C|\Phi\rangle$
	–	$\langle\Phi_{\tilde{j}}^{\tilde{b}}|(V_{N,C}T_{3C})_C|\Phi\rangle$
	–	$\langle\Phi_j^b|(V_{N,A}T_{1A}T_{2A})_C|\Phi\rangle$
	–	$\langle\Phi_j^b|(V_{N,B}T_{1A}T_{2B})_C|\Phi\rangle$
	–	$\langle\Phi_{\tilde{j}}^{b}|(V_{N,B}T_{1B}T_{2A})_C|\Phi\rangle$
	–	$\langle\Phi_j^b|(V_{N,C}T_{1B}T_{2B})_C|\Phi\rangle$
	–	$\frac{1}{6}\langle\Phi_j^b|(V_{N,A}T_{1A}T_{1A}T_{1A})_C|\Phi\rangle$
	–	$\frac{1}{2}\langle\Phi_j^b|(V_{N,B}T_{1A}T_{1A}T_{1B})_C|\Phi\rangle$

Table 4.11: Spin-integrated and factorized terms resulting from the semi-disconnected diagrams resulting from Eq. (4.39) projected onto doubly excited determinants of the α/β type.

	$(t_{a\tilde{b}}^{i\tilde{j}} + t_a^i t_{\tilde{b}}^{\tilde{j}})\times$	$t_a^i \times$	$t_{\tilde{b}}^{\tilde{j}} \times$
zero-body	$\langle\Phi|(V_{N,A}T_{2A})_C|\Phi\rangle$	–	–
	$\langle\Phi|(V_{N,B}T_{2B})_C|\Phi\rangle$	–	–
	$\langle\Phi|(V_{N,C}T_{2C})_C|\Phi\rangle$	–	–
	$\frac{1}{2}\langle\Phi|(V_{N,A}T_{1A}T_{1A})_C|\Phi\rangle$	–	–
	$\langle\Phi|(V_{N,B}T_{1A}T_{1B})_C|\Phi\rangle$	–	–
	$\frac{1}{2}\langle\Phi|(V_{N,C}T_{1B}T_{1B})_C|\Phi\rangle$	–	–
one-body	–	$\langle\Phi_{\tilde{j}}^{\tilde{b}}|(V_{N,A}T_{3B})_C|\Phi\rangle$	$\langle\Phi_i^a|(V_{N,A}T_{3A})_C|\Phi\rangle$
	–	$\langle\Phi_{\tilde{j}}^{\tilde{b}}|(V_{N,B}T_{3C})_C|\Phi\rangle$	$\langle\Phi_i^a|(V_{N,B}T_{3B})_C|\Phi\rangle$
	–	$\langle\Phi_{\tilde{j}}^{\tilde{b}}|(V_{N,C}T_{3D})_C|\Phi\rangle$	$\langle\Phi_i^a|(V_{N,C}T_{3C})_C|\Phi\rangle$
	–	$\langle\Phi_{\tilde{j}}^{\tilde{b}}|(V_{N,B}T_{1B}T_{2B})_C|\Phi\rangle$	$\langle\Phi_i^a|(V_{N,A}T_{1A}T_{2A})_C|\Phi\rangle$
	–	$\langle\Phi_{\tilde{j}}^{\tilde{b}}|(V_{N,C}T_{1B}T_{2C})_C|\Phi\rangle$	$\langle\Phi_i^a|(V_{N,B}T_{1A}T_{2B})_C|\Phi\rangle$
	–	$\frac{1}{2}\langle\Phi_{\tilde{j}}^{\tilde{b}}|(V_{N,B}T_{1A}T_{1B}T_{1B})_C|\Phi\rangle$	$\frac{1}{6}\langle\Phi_i^a|(V_{N,A}T_{1A}T_{1A}T_{1A})_C|\Phi\rangle$
	–	$\frac{1}{6}\langle\Phi_{\tilde{j}}^{\tilde{b}}|(V_{N,C}T_{1B}T_{1B}T_{1B})_C|\Phi\rangle$	$\frac{1}{2}\langle\Phi_i^a|(V_{N,B}T_{1A}T_{1A}T_{1B})_C|\Phi\rangle$

119

Table 4.12: Spin-integrated and factorized terms resulting from the semi-disconnected diagrams resulting from Eq. (4.39) projected onto doubly excited determinants of the β/β type.

	$(t^{\tilde{i}\tilde{j}}_{\tilde{a}\tilde{b}} + \mathscr{A}_{\tilde{i}\tilde{j}} t^{\tilde{i}}_{\tilde{a}} t^{\tilde{j}}_{\tilde{b}}) \times$	$\mathscr{A}^{\tilde{a}\tilde{b}}_{\tilde{i}\tilde{j}} t^{\tilde{i}}_{\tilde{a}} \times$
zero-body	$\langle\Phi|(V_{N,A}T_{2A})_C|\Phi\rangle$	–
	$\langle\Phi|(V_{N,B}T_{2B})_C|\Phi\rangle$	–
	$\langle\Phi|(V_{N,C}T_{2C})_C|\Phi\rangle$	–
	$\frac{1}{2}\langle\Phi|(V_{N,A}T_{1A}T_{1A})_C|\Phi\rangle$	–
	$\langle\Phi|(V_{N,B}T_{1A}T_{1B})_C|\Phi\rangle$	–
	$\frac{1}{2}\langle\Phi|(V_{N,C}T_{1B}T_{1B})_C|\Phi\rangle$	–
one-body	–	$\langle\Phi^{\tilde{b}}_{\tilde{j}}|(V_{N,A}T_{3B})_C|\Phi\rangle$
	–	$\langle\Phi^{\tilde{b}}_{\tilde{j}}|(V_{N,B}T_{3C})_C|\Phi\rangle$
	–	$\langle\Phi^{\tilde{b}}_{\tilde{j}}|(V_{N,C}T_{3D})_C|\Phi\rangle$
	–	$\langle\Phi^{\tilde{b}}_{\tilde{j}}|(V_{N,C}T_{1B}T_{2C})_C|\Phi\rangle$
	–	$\langle\Phi^{\tilde{b}}_{\tilde{j}}|(V_{N,B}T_{1B}T_{2B})_C|\Phi\rangle$
	–	$\langle\Phi^{\tilde{b}}_{\tilde{j}}|(V_{N,B}T_{1A}T_{2C})_C|\Phi\rangle$
	–	$\langle\Phi^{\tilde{b}}_{\tilde{j}}|(V_{N,A}T_{1A}T_{2B})_C|\Phi\rangle$
	–	$\frac{1}{2}\langle\Phi^{\tilde{b}}_{\tilde{j}}|(V_{N,B}T_{1A}T_{1B}T_{1B})_C|\Phi\rangle$
	–	$\frac{1}{6}\langle\Phi^{\tilde{b}}_{\tilde{j}}|(V_{N,C}T_{1B}T_{1B}T_{1B})_C|\Phi\rangle$

4.10, 4.11, and 4.12 were computed by taking outer products of the disconnected pieces.

It is worth noting that in our considerations we have only optimized the quadruply excited components. However, the treatment described in this section can be applied to any term where T_3 clusters also appear.

CHAPTER 5

CONCLUSIONS AND FUTURE OUTLOOK

In this Ph.D. dissertation work, we have given a brief introduction to the state of affairs in the challenges and solutions faced by the area of QC. We have also discussed the fundamentals of CC theory, including the standard SRCC approaches, such as CCSD, CCSDT, CCSDTQ, etc., their active-space extensions, their application to excited electronic states via the EOM formalism, and the MM corrections of CC equations. With this, we presented the modern CC($P;Q$) framework, resulting in methods like CC(t;3), CC(t,q;3), CC(t,q,;3,4), and CC(q,4), which allows us to extend any of these ideas to conventional and unconventional truncations and definitions of the many-body expansion characterizing the cluster operator, T.

This introduction allowed us to introduce the main topic of this thesis, which is the merger of the CC($P;Q$) methodology with the FCIQMC methods of Professor Alavi *et al.*, leading to the novel semi-stochastic CC($P;Q$) approaches developed during this Ph.D. work [202, 203]. Their application to ground and excited electronic states, was discussed during Chapter 3, using the examples provided by the bond breaking of the F_2, simultaneous bond stretching of H_2O, automerization reaction of cyclobutadiene, and potential energy curves corresponding to various excited electronic states of the CH^+ ion. The results show how powerful the semi-stochastic approaches are in recovering the energies of higher-order SRCC theories, such as CCSDT, CCSDTQ, and EOMCCSDT, at a fraction of the computational cost.

Furthermore, in this dissertation, we have presented an alternative way of exploiting the information produced during *i*-FCIQMC simulations, this time aiming toward the recovery of exact, FCI, ground-state energies [204]. We have conceived a method, reminiscent of the externally corrected approaches historically used to correct CC calculations by using particularly robust non-CC wave function sources, where we cluster analyze

the *i*-FCIQMC wave function, to extract triply and quadruply cluster components that we subsequently utilize to solve CCSD-like equations. The performance of this second methodology introduced in this Ph.D. thesis, has been tested by calculating the ground-state energies of the simultaneous bond stretching of H_2O, as shown in Chapter 3. The results of this study show that the CAD-FCIQMC methodology is capable of accelerating the convergence of *i*-FCIQMC simulations, by producing much more reliable energetics, with less noise, even at the early stages of the *i*-FCIQMC simulations.

Finally, in this study, we have also presented the implementation strategies and computational techniques employed to take advantage of the reduced-dimensionality of the semi-stochastic CC(P;Q) and CAD-FCIQMC methods relative to the full treatment of the corresponding CCSDT, CCSDTQ, or EOMCCSDT parents, in the case of our CC(P;Q) approaches, and to FCI when considering the CAD-FCIQMC method. We have shown that sparse and stochastically determined and spin-integrated P spaces can be taken advantage by performing the corresponding tensor contractions using matrix elements of the similarity transformed Hamiltonian in the N-electron basis given by excited Slater determinants, and that these elements can be computed on-the-fly, without resorting to the full construction of the corresponding matrices. We have also shown that the cluster-analysis behind the CAD-FCIQMC calculations can be optimized by directly computing the relevant projections of the memory- and CPU-demanding T_4 component onto doubly excited determinants, bypassing the need for the construction of the corresponding T_4 matrices, since *i*-FCIQMC simulations provide quite sparse C_4 coefficients.

Further work is already being undertaken by members of the Piecuch's group, who are extending the CC(P;Q) methodology to excited states and open-shell systems. Also, the optimizations described in this work are still underway and need to be extended to all stages of the CC(P;Q) calculation, namely the EOM and left-CC parts of the calculation.

Concerning the CAD-FCIQMC method, new studies are in preparation detailing the

use of cluster analysis to adaptive CI methods, such as CIPSI an others. In addition to this, more work is still required for optimizing the T_3 cluster component in our software implementation of CAD-FCIQMC, by exploiting on-the-fly projections of the connected cluster form of the Schrödinger equation, containing T_3 clusters, onto singly and doubly excited determinants.

This work has proven the importance of exploring novel ways of exploiting wave-function-based QC methodologies that take advantage of the strong sparsity of the many-body Hilbert space. It is our hope that this work encourages further studies of more general semi-stochastic CC(P;Q) approaches that can include any subset of the Hilbert space without restrictions on the excitation rank, or even the exploration of novel ways of sampling the Hilbert space. In the case of the CAD-FCIQMC methodology, there is plenty of space ahead to study other ways of incorporating its features in accelerating the convergence within i-FCIQMC or other CI or CC methodologies. For instance, one can envision new methods were the resulting components of the cluster operator in CAD-FCIQMC calculations can be reused to extrapolate i-FCIQMC wave functions, during the stochastic propagation along the imaginary time axis.

APPENDIX

A.1 Spin-Integrated and Rank-Conserving Hugenholtz and Brandow Diagrams Corresponding to the Similarity Transformed Hamiltonian with up to Doubly Excited Cluster Component

Table .1: Brandow diagrams and the corresponding algebraic expressions representing the spin-integrated one-body components of the term $F_N T_1$ of $\bar{H}_1$. The absence/presence of tilde symbols on top of indices defines the α/β spin of the corresponding particle or hole line, respectively.

$$f_j^e t_e^i \, N[a^j a_i] \qquad -f_m^b t_a^m \, N[a^a a_b] \qquad f_{\tilde{j}}^{\tilde{e}} t_{\tilde{e}}^{\tilde{i}} \, N[a^{\tilde{j}} a_{\tilde{i}}] \qquad -f_{\tilde{m}}^{\tilde{b}} t_{\tilde{a}}^{\tilde{m}} \, N[a^{\tilde{a}} a_{\tilde{b}}]$$

Table .2: Hugenholtz and Brandow diagrams with the corresponding algebraic expressions representing the one-body components of the term $V_N T_1$ of $\bar{H}_1$. The absence/presence of tilde symbols on top of indices defined the α/β spin of the corresponding particle or hole line, respectively.

$v_{jm}^{ie} t_e^m \, N[a^j a_i]$ $v_{am}^{be} t_e^m \, N[a^a a_b]$ $v_{\tilde{j}\tilde{m}}^{\tilde{i}\tilde{e}} t_{\tilde{e}}^{\tilde{m}} \, N[a^{\tilde{j}} a_{\tilde{i}}]$ $v_{\tilde{a}\tilde{m}}^{\tilde{b}\tilde{e}} t_{\tilde{e}}^{\tilde{m}} \, N[a^a a_{\tilde{b}}]$

$v_{j\tilde{m}}^{i\tilde{e}} t_{\tilde{e}}^{\tilde{m}} \, N[a^j a_i]$ $v_{am}^{be} t_e^m \, N[a^a a_b]$ $v_{\tilde{j}m}^{\tilde{i}e} t_e^m \, N[a^{\tilde{j}} a_{\tilde{i}}]$ $v_{\tilde{a}m}^{\tilde{b}e} t_e^m \, N[a^{\tilde{a}} a_{\tilde{b}}]$

126

Table .3: Brandow diagrams and the corresponding algebraic expressions representing the spin-integrated one-body components of the term $V_N T_2$ of $\bar{H}_1$. The absence/presence of tilde symbols on top of indices defines the α/β spin of the corresponding particle or hole line, respectively.

$$\frac{1}{2}v_{jm}^{fe}t_{fe}^{im}\,N[a^j a_i] \qquad \frac{1}{2}v_{\tilde{j}\tilde{m}}^{\tilde{f}\tilde{e}}t_{\tilde{f}\tilde{e}}^{\tilde{i}\tilde{m}}\,N[a^{\tilde{j}} a_{\tilde{i}}] \qquad v_{jm}^{f\tilde{e}}t_{f\tilde{e}}^{i\tilde{m}}\,N[a^j a_i] \qquad v_{\tilde{j}m}^{\tilde{f}e}t_{\tilde{f}e}^{\tilde{i}m}\,N[a^{\tilde{j}} a_{\tilde{i}}]$$

$$\frac{1}{2}v_{nm}^{ae}t_{be}^{nm}\,N[a^b a_a] \qquad \frac{1}{2}v_{\tilde{n}\tilde{m}}^{\tilde{a}\tilde{e}}t_{\tilde{b}\tilde{e}}^{\tilde{n}\tilde{m}}\,N[a^b a_{\tilde{a}}] \qquad v_{nm}^{a\tilde{e}}t_{b\tilde{e}}^{n\tilde{m}}\,N[a^b a_a] \qquad v_{\tilde{n}m}^{\tilde{a}e}t_{\tilde{b}e}^{\tilde{n}m}\,N[a^{\tilde{b}} a_{\tilde{a}}]$$

Table .4: Hugenholtz and Brandow diagrams with the corresponding algebraic expressions representing the two-body components of the term $V_N T_1$ of $\bar{H}_2$. The absence/presence of tilde symbols on top of indices defined the α/β spin of the corresponding particle or hole line, respectively.

$\frac{1}{4}\mathscr{A}^{ij}v^{ej}_{kl}t^i_e\,N[a^k a_i a^l a_j]$

$\frac{1}{4}\mathscr{A}^{\tilde{i}\tilde{j}}v^{\tilde{e}\tilde{j}}_{\tilde{k}\tilde{l}}t^{\tilde{i}}_{\tilde{e}}\,N[a^{\tilde{k}} a_{\tilde{i}} a^{\tilde{l}} a_{\tilde{j}}]$

$v^{e\tilde{j}}_{k\tilde{l}}t^i_e\,N[a^k a_i a^{\tilde{l}} a_{\tilde{j}}]$

$v^{\tilde{e}j}_{\tilde{k}l}t^{\tilde{i}}_{\tilde{e}}\,N[a^{\tilde{k}} a_{\tilde{i}} a^l a_j]$

$v^{eb}_{ja}t^i_e\,N[a^j a_i a^a a_b]$

$v^{\tilde{e}\tilde{b}}_{\tilde{j}\tilde{a}}t^{\tilde{i}}_{\tilde{e}}\,N[a^{\tilde{j}} a_{\tilde{i}} a^{\tilde{a}} a_{\tilde{b}}]$

$v^{e\tilde{b}}_{j\tilde{a}}t^i_e\,N[a^j a_i a^{\tilde{a}} a_{\tilde{b}}]$

$v^{\tilde{e}b}_{\tilde{j}a}t^{\tilde{i}}_{\tilde{e}}\,N[a^{\tilde{j}} a_{\tilde{i}} a^a a_b]$

128

$$v^{bi}_{mj} t^m_a\, N[a^a a_b a^j a_i] \qquad v^{\tilde b \tilde i}_{\tilde m \tilde j} t^{\tilde m}_{\tilde a}\, N[a^{\tilde a} a_{\tilde b} a^{\tilde j} a_{\tilde i}] \qquad v^{b\tilde i}_{mj} t^m_a\, N[a^a a_b a^{\tilde j} a_{\tilde i}] \qquad v^{\tilde b i}_{\tilde m j} t^{\tilde m}_{\tilde a}\, N[a^{\tilde a} a_{\tilde b} a^j a_i]$$

$$\tfrac{1}{4}\mathscr{A}_{ab} v^{cd}_{mb} t^m_a\, N[a^a a_c a^b a_d] \qquad \tfrac{1}{4}\mathscr{A}_{\tilde a \tilde b} v^{\tilde c \tilde d}_{\tilde m \tilde b} t^{\tilde m}_{\tilde a}\, N[a^{\tilde a} a_{\tilde c} a^{\tilde b} a_{\tilde d}] \qquad v^{c\tilde d}_{mb} t^m_a\, N[a^a a_c a^{\tilde b} a_{\tilde d}] \qquad v^{\tilde c d}_{\tilde m b} t^{\tilde m}_{\tilde a}\, N[a^{\tilde a} a_{\tilde c} a^b a_d]$$

Table .5: Hugenholtz and Brandow diagrams with the corresponding algebraic expressions representing the two-body components of the term $V_N T_2$ of $\bar{H}_2$. The absence/presence of tilde symbols on top of indices defined the α/β spin of the corresponding particle or hole line, respectively.

$$\frac{1}{8} v^{ef}_{kl} t^{ij}_{ef} \, N[a^k a_i a^l a_j] \qquad \frac{1}{8} v^{\tilde{e}\tilde{f}}_{\tilde{k}\tilde{l}} t^{\tilde{i}\tilde{j}}_{\tilde{e}\tilde{f}} \, N[a^{\tilde{k}} a_{\tilde{i}} a^{\tilde{l}} a_{\tilde{j}}] \qquad v^{e\tilde{f}}_{k\tilde{l}} t^{i\tilde{j}}_{e\tilde{f}} \, N[a^k a_i a^{\tilde{l}} a_{\tilde{j}}]$$

130

$$v_{fj}^{eb} t_{ae}^{if} N[a^a a_i a^b a_j]$$

$$v_{\tilde{f}\tilde{j}}^{\tilde{e}\tilde{b}} t_{\tilde{a}\tilde{e}}^{\tilde{i}\tilde{f}} N[a^{\tilde{a}} a_{\tilde{i}} a^{\tilde{b}} a_{\tilde{j}}]$$

$$v_{j\tilde{f}}^{b\tilde{e}} t_{a\tilde{e}}^{i\tilde{f}} N[a^a a_i a^b a_j]$$

$$v^{\tilde{b}e}_{\tilde{j}f}t^{\tilde{i}f}_{\tilde{a}e}\,N[a^{\tilde{a}}a_{\tilde{i}}a^{\tilde{b}}a_{\tilde{j}}]$$

$$v^{b\tilde{e}}_{m\tilde{j}}t^{m\tilde{i}}_{a\tilde{e}}\,N[a^a a_b a^{\tilde{j}}a_{\tilde{i}}]$$

$$v^{\tilde{b}e}_{\tilde{m}j}t^{\tilde{m}i}_{\tilde{a}e}\,N[a^{\tilde{a}}a_{\tilde{b}}a^j a_i]$$

$$\frac{1}{8}v_{kl}^{ef}t_{ef}^{ij}\,N[a^k a_i a^l a_j] \qquad \frac{1}{8}v_{\tilde{k}\tilde{l}}^{\tilde{e}\tilde{f}}t_{\tilde{e}\tilde{f}}^{\tilde{i}\tilde{j}}\,N[a^{\tilde{k}} a_{\tilde{i}} a^{\tilde{l}} a_{\tilde{j}}] \qquad v_{k\tilde{l}}^{e\tilde{f}}t_{e\tilde{f}}^{i\tilde{j}}\,N[a^k a_i a^{\tilde{l}} a_{\tilde{j}}]$$

133

Table .6: Hugenholtz and Brandow diagrams with the corresponding algebraic expressions representing the two-body components of the term $V_N T_1 T_1$ of $\bar{H}_2$. The absence/presence of tilde symbols on top of indices defined the α/β spin of the corresponding particle or hole line, respectively.

$$v^{be}_{mj} t^m_a t^i_e \, N[a^a a_b a^j a_i] \quad v^{\tilde{b}\tilde{e}}_{\tilde{m}j} t^{\tilde{m}}_{\tilde{a}} t^{\tilde{i}}_{\tilde{e}} a^{\tilde{a}} \, N[a_{\tilde{b}} a^{\tilde{j}} a_{\tilde{i}}] \quad v^{b\tilde{e}}_{mj} t^m_a t^{\tilde{i}}_{\tilde{e}} \, N[a^a a_b a^{\tilde{j}} a_{\tilde{i}}] \quad v^{\tilde{b}e}_{\tilde{m}j} t^{\tilde{m}}_{\tilde{a}} t^i_e \, N[a^{\tilde{a}} a_{\tilde{b}} a^j a_i]$$

134

$$\frac{1}{4}v_{kl}^{ef}\, t_e^i t_f^j\, N[a^k a_i a^l a_j] \qquad \frac{1}{4}v_{\tilde{k}\tilde{l}}^{\tilde{e}\tilde{f}}\, t_{\tilde{e}}^{\tilde{i}} t_{\tilde{f}}^{\tilde{j}}\, N[a^{\tilde{k}} a_{\tilde{i}} a^{\tilde{l}} a_{\tilde{j}}] \qquad v_{kl}^{e\tilde{f}}\, t_e^i t_{\tilde{f}}^{\tilde{j}}\, N[a^k a_i a^{\tilde{l}} a_{\tilde{j}}]$$

$$\frac{1}{4}v_{mn}^{cd}\, t_a^m t_b^n\, N[a^a a_c a^b a_d] \qquad \frac{1}{4}v_{\tilde{m}\tilde{n}}^{\tilde{c}\tilde{d}}\, t_{\tilde{a}}^{\tilde{m}} t_{\tilde{b}}^{\tilde{n}}\, N[a^{\tilde{a}} a_{\tilde{c}} a^{\tilde{b}} a_{\tilde{d}}] \qquad v_{m\tilde{n}}^{c\tilde{d}}\, t_a^m t_{\tilde{b}}^{\tilde{n}}\, N[a^a a_c a^{\tilde{b}} a_{\tilde{d}}]$$

Table .7: Hugenholtz and Brandow diagrams corresponding to the particle conserving three-body components $V_N T_2$ within the $\bar{H}$ operator including their corresponding algebraic expressions. The absence/presence of tilde symbols on top of indices defined the α/β spin of the corresponding particle or hole line, respectively.

$$\tfrac{1}{4} v^{eb}_{kl} t^{ij}_{ae} N[a^a a^k a^l a_i a_j a_b] \quad -\tfrac{1}{4} v^{cd}_{mj} t^{im}_{ab} N[a^a a^b a^j a_i a_c a_d] \quad \tfrac{1}{4} v^{\tilde{e}\tilde{b}}_{\tilde{k}\tilde{l}} t^{\tilde{i}\tilde{j}}_{\tilde{a}\tilde{e}} N[a^{\tilde{a}} a^{\tilde{k}} a^{\tilde{l}} a_{\tilde{i}} a_{\tilde{j}} a_{\tilde{b}}] \quad -\tfrac{1}{4} v^{\tilde{c}\tilde{d}}_{\tilde{m}\tilde{j}} t^{\tilde{i}\tilde{m}}_{\tilde{a}\tilde{b}} N[a^{\tilde{a}} a^{\tilde{b}} a^{\tilde{j}} a_{\tilde{i}} a_{\tilde{c}} a_{\tilde{d}}]$$

$$v^{\tilde{e}b}_{\tilde{k}l} t^{i\tilde{j}}_{a\tilde{e}} N[a^a a^{\tilde{k}} a^l a_i a_{\tilde{j}} a_b] \quad -v^{\tilde{c}d}_{\tilde{m}j} t^{i\tilde{m}}_{ab} N[a^a a^{\tilde{b}} a^j a_i a_{\tilde{c}} a_d] \quad v^{e\tilde{b}}_{k\tilde{l}} t^{\tilde{i}j}_{\tilde{a}e} N[a^{\tilde{a}} a^k a^{\tilde{l}} a_{\tilde{i}} a_j a_{\tilde{b}}] \quad -v^{c\tilde{d}}_{m\tilde{j}} t^{\tilde{i}m}_{\tilde{a}b} N[a^{\tilde{a}} a^b a^{\tilde{j}} a_{\tilde{i}} a_c a_{\tilde{d}}]$$

136

$$\frac{1}{2}v_{kl}^{e\tilde{b}}t_{\tilde{a}e}^{ij}\,N[a^aa^ka^{\tilde{l}}a_ia_ja_{\tilde{b}}] \quad -\frac{1}{2}v_{mj}^{c\tilde{d}}t_{ab}^{im}\,N[a^aa^ba^{\tilde{j}}a_ia_ca_{\tilde{d}}] \quad \frac{1}{2}v_{\tilde{k}\tilde{l}}^{\tilde{e}\tilde{b}}t_{a\tilde{e}}^{i\tilde{j}}\,N[a^aa^{\tilde{k}}a^{\tilde{l}}a_ia_{\tilde{j}}a_{\tilde{b}}] \quad -\frac{1}{2}v_{\tilde{m}j}^{\tilde{c}\tilde{d}}t_{a\tilde{b}}^{i\tilde{m}}\,N[a^aa^{\tilde{b}}a^{\tilde{j}}a_ia_{\tilde{c}}a_{\tilde{d}}]$$

$$\frac{1}{2}v_{kl}^{eb}t_{\tilde{a}e}^{\tilde{i}j}\,N[a^{\tilde{a}}a^ka^la_{\tilde{i}}a_ja_b] \quad -\frac{1}{2}v_{mj}^{cd}t_{\tilde{a}b}^{\tilde{i}m}\,N[a^{\tilde{a}}a^ba^ja_{\tilde{i}}a_ca_d] \quad \frac{1}{2}v_{\tilde{k}l}^{\tilde{e}b}t_{\tilde{a}\tilde{e}}^{\tilde{i}\tilde{j}}\,N[a^{\tilde{a}}a^{\tilde{k}}a^la_{\tilde{i}}a_{\tilde{j}}a_b] \quad -\frac{1}{2}v_{\tilde{m}j}^{\tilde{c}d}t_{\tilde{a}b}^{\tilde{i}\tilde{m}}\,N[a^{\tilde{a}}a^{\tilde{b}}a^ja_{\tilde{i}}a_{\tilde{c}}a_d]$$

A.2 Spin-Integrated Matrix Elements of the Three-Body Component of the Similarity Transformed Hamiltonian

A.2.1 Diagonal Terms

$$
\begin{aligned}
\langle \Phi_{ijk}^{abc} | \bar{H}_3 | \Phi_{ijk}^{abc} \rangle = & -\bar{h}_{aij}^{iaj} - \bar{h}_{aik}^{iak} - \bar{h}_{ajk}^{jak} \\
& - \bar{h}_{bij}^{ibj} - \bar{h}_{bik}^{ibk} - \bar{h}_{bjk}^{jbk} \\
& - \bar{h}_{cij}^{icj} - \bar{h}_{cik}^{ick} - \bar{h}_{cjk}^{jck} \\
& + \bar{h}_{aib}^{iab} + \bar{h}_{aic}^{iac} + \bar{h}_{bic}^{ibc} \\
& + \bar{h}_{ajb}^{jab} + \bar{h}_{ajc}^{jac} + \bar{h}_{bjc}^{jbc} \\
& + \bar{h}_{akb}^{kab} + \bar{h}_{akc}^{kac} + \bar{h}_{bkc}^{kbc}
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}} | \bar{H}_3 | \Phi_{ij\tilde{k}}^{ab\tilde{c}} \rangle = & -\bar{h}_{aij}^{iaj} - \bar{h}_{ai\tilde{k}}^{ia\tilde{k}} - \bar{h}_{aj\tilde{k}}^{ja\tilde{k}} \\
& - \bar{h}_{bij}^{ibj} - \bar{h}_{bi\tilde{k}}^{ib\tilde{k}} - \bar{h}_{bj\tilde{k}}^{jb\tilde{k}} \\
& - \bar{h}_{\tilde{c}i\tilde{k}}^{i\tilde{c}\tilde{k}} - \bar{h}_{\tilde{c}j\tilde{k}}^{j\tilde{c}\tilde{k}} \\
& + \bar{h}_{aib}^{iab} + \bar{h}_{ai\tilde{c}}^{ia\tilde{c}} + \bar{h}_{bi\tilde{c}}^{ib\tilde{c}} \\
& + \bar{h}_{ajb}^{jab} + \bar{h}_{aj\tilde{c}}^{ja\tilde{c}} + \bar{h}_{bj\tilde{c}}^{jb\tilde{c}} \\
& + \bar{h}_{a\tilde{k}\tilde{c}}^{\tilde{k}a\tilde{c}} + \bar{h}_{b\tilde{k}\tilde{c}}^{\tilde{k}b\tilde{c}}
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
\langle \Phi_{i\tilde{j}\tilde{k}}^{a\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{i\tilde{j}\tilde{k}}^{a\tilde{b}\tilde{c}} \rangle = & -\bar{h}_{ai\tilde{j}}^{ia\tilde{j}} - \bar{h}_{ai\tilde{k}}^{ia\tilde{k}} \\
& - \bar{h}_{\tilde{b}i\tilde{j}}^{i\tilde{b}\tilde{j}} - \bar{h}_{\tilde{b}i\tilde{k}}^{i\tilde{b}\tilde{k}} - \bar{h}_{\tilde{b}\tilde{j}\tilde{k}}^{\tilde{j}\tilde{b}\tilde{k}} \\
& - \bar{h}_{\tilde{c}i\tilde{j}}^{i\tilde{c}\tilde{j}} - \bar{h}_{\tilde{c}i\tilde{k}}^{i\tilde{c}\tilde{k}} - \bar{h}_{\tilde{c}\tilde{j}\tilde{k}}^{\tilde{j}\tilde{c}\tilde{j}} \\
& + \bar{h}_{ai\tilde{b}}^{ia\tilde{b}} + \bar{h}_{ai\tilde{c}}^{ia\tilde{c}} \\
& + \bar{h}_{a\tilde{j}\tilde{b}}^{\tilde{j}a\tilde{b}} + \bar{h}_{a\tilde{j}\tilde{c}}^{\tilde{j}a\tilde{c}} + \bar{h}_{\tilde{b}\tilde{j}\tilde{c}}^{\tilde{j}\tilde{b}\tilde{c}} \\
& + \bar{h}_{a\tilde{k}\tilde{b}}^{\tilde{k}a\tilde{b}} + \bar{h}_{a\tilde{k}\tilde{c}}^{\tilde{k}a\tilde{c}} + \bar{h}_{\tilde{b}\tilde{k}\tilde{c}}^{\tilde{k}\tilde{b}\tilde{c}}
\end{aligned}
\tag{3}
$$

$$\langle \Phi_{\tilde{i}\tilde{j}\tilde{k}}^{\tilde{a}\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{\tilde{i}\tilde{j}\tilde{k}}^{\tilde{a}\tilde{b}\tilde{c}} \rangle = - \bar{h}_{\tilde{a}\tilde{i}j}^{\tilde{i}\tilde{a}\tilde{j}} - \bar{h}_{\tilde{a}\tilde{i}k}^{\tilde{i}\tilde{a}\tilde{k}} - \bar{h}_{\tilde{a}\tilde{j}k}^{\tilde{j}\tilde{a}\tilde{k}}$$

$$- \bar{h}_{\tilde{b}\tilde{i}j}^{\tilde{i}\tilde{b}\tilde{j}} - \bar{h}_{\tilde{b}\tilde{i}k}^{\tilde{i}\tilde{b}\tilde{k}} - \bar{h}_{\tilde{b}\tilde{j}k}^{\tilde{j}\tilde{b}\tilde{k}}$$

$$- \bar{h}_{\tilde{c}\tilde{i}j}^{\tilde{i}\tilde{c}\tilde{j}} - \bar{h}_{\tilde{c}\tilde{i}k}^{\tilde{i}\tilde{c}\tilde{k}} - \bar{h}_{\tilde{c}\tilde{j}k}^{\tilde{j}\tilde{c}\tilde{k}}$$

$$+ \bar{h}_{\tilde{a}\tilde{i}b}^{\tilde{i}\tilde{a}\tilde{b}} + \bar{h}_{\tilde{a}\tilde{i}c}^{\tilde{i}\tilde{a}\tilde{c}} + \bar{h}_{\tilde{b}\tilde{i}c}^{\tilde{i}\tilde{b}\tilde{c}}$$

$$+ \bar{h}_{\tilde{a}\tilde{j}b}^{\tilde{j}\tilde{a}\tilde{b}} + \bar{h}_{\tilde{a}\tilde{j}c}^{\tilde{j}\tilde{a}\tilde{c}} + \bar{h}_{\tilde{b}\tilde{j}c}^{\tilde{j}\tilde{b}\tilde{c}}$$

$$+ \bar{h}_{\tilde{a}\tilde{k}b}^{\tilde{k}\tilde{a}\tilde{b}} + \bar{h}_{\tilde{a}\tilde{k}c}^{\tilde{k}\tilde{a}\tilde{c}} + \bar{h}_{\tilde{b}\tilde{k}c}^{\tilde{k}\tilde{b}\tilde{c}}$$

$$(4)$$

A.2.2 Off-diagonal Elements with one Spin-Orbital Difference

$$\langle\Phi^{abc}_{ijk}|\bar{H}_3|\Phi^{abc'}_{ijk}\rangle = -\bar{h}^{ic'j}_{cij} - \bar{h}^{ic'k}_{cik} - \bar{h}^{jc'k}_{cjk}$$
$$+ \bar{h}^{iac'}_{aic} + \bar{h}^{ibc'}_{bic}$$
$$+ \bar{h}^{jac'}_{ajc} + \bar{h}^{jbc'}_{bjc}$$
$$+ \bar{h}^{kac'}_{akc} + \bar{h}^{kbc'}_{bkc}$$

(5)

$$\langle\Phi^{ab\tilde{c}}_{ij\tilde{k}}|\bar{H}_3|\Phi^{ab'\tilde{c}}_{ij\tilde{k}}\rangle = -\bar{h}^{ib'j}_{bij} - \bar{h}^{ib'\tilde{k}}_{bi\tilde{k}} - \bar{h}^{jb'\tilde{k}}_{bj\tilde{k}}$$
$$+ \bar{h}^{iab'}_{aib} + \bar{h}^{ib'\tilde{c}}_{bi\tilde{c}}$$
$$+ \bar{h}^{jab'}_{ajb} + \bar{h}^{jb'\tilde{c}}_{bj\tilde{c}}$$
$$+ \bar{h}^{\tilde{k}b'\tilde{c}}_{b\tilde{k}\tilde{c}}$$

(6)

$$\langle\Phi^{a\tilde{b}\tilde{c}}_{i\tilde{j}\tilde{k}}|\bar{H}_3|\Phi^{a'\tilde{b}\tilde{c}}_{i\tilde{j}\tilde{k}}\rangle = -\bar{h}^{ia'\tilde{j}}_{ai\tilde{j}} - \bar{h}^{ia'\tilde{k}}_{ai\tilde{k}}$$
$$+ \bar{h}^{ia'\tilde{b}}_{ai\tilde{b}} + \bar{h}^{ia'\tilde{c}}_{ai\tilde{c}}$$
$$+ \bar{h}^{\tilde{j}a'\tilde{b}}_{a\tilde{j}\tilde{b}} + \bar{h}^{\tilde{j}a'\tilde{c}}_{a\tilde{j}\tilde{c}}$$
$$+ \bar{h}^{\tilde{k}a'\tilde{b}}_{a\tilde{k}\tilde{b}} + \bar{h}^{\tilde{k}a'\tilde{c}}_{a\tilde{k}\tilde{c}}$$

(7)

$$\langle\Phi^{ab\tilde{c}}_{ij\tilde{k}}|\bar{H}_3|\Phi^{ab\tilde{c'}}_{ij\tilde{k}}\rangle = -\bar{h}^{i\tilde{c'}\tilde{k}}_{\tilde{c}ik} - \bar{h}^{j\tilde{c'}\tilde{k}}_{\tilde{c}jk}$$
$$+ \bar{h}^{ia\tilde{c'}}_{ai\tilde{c}} + \bar{h}^{ib\tilde{c'}}_{bi\tilde{c}}$$
$$+ \bar{h}^{ja\tilde{c'}}_{aj\tilde{c}} + \bar{h}^{jb\tilde{c'}}_{bj\tilde{c}}$$
$$+ \bar{h}^{\tilde{k}a\tilde{c'}}_{a\tilde{k}\tilde{c}} + \bar{h}^{\tilde{k}b\tilde{c'}}_{b\tilde{k}\tilde{c}}$$

(8)

$$\langle\Phi^{a\tilde{b}\tilde{c}}_{i\tilde{j}\tilde{k}}|\bar{H}_3|\Phi^{a\tilde{b}\tilde{c'}}_{i\tilde{j}\tilde{k}}\rangle = -\bar{h}^{i\tilde{c'}\tilde{k}}_{\tilde{c}i\tilde{k}} - \bar{h}^{i\tilde{c'}\tilde{j}}_{\tilde{c}i\tilde{j}} - \bar{h}^{\tilde{j}\tilde{c'}\tilde{k}}_{\tilde{c}\tilde{j}\tilde{k}}$$
$$+ \bar{h}^{ia\tilde{c'}}_{ai\tilde{c}}$$
$$+ \bar{h}^{\tilde{j}a\tilde{c'}}_{a\tilde{j}\tilde{c}} + \bar{h}^{\tilde{j}\tilde{b}\tilde{c'}}_{\tilde{b}\tilde{j}\tilde{c}}$$
$$+ \bar{h}^{\tilde{k}a\tilde{c'}}_{a\tilde{k}\tilde{c}} + \bar{h}^{\tilde{k}\tilde{b}\tilde{c'}}_{\tilde{b}\tilde{k}\tilde{c}}$$

(9)

140

$$\langle \Phi_{\tilde{i}\tilde{j}\tilde{k}}^{\tilde{a}\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{\tilde{i}\tilde{j}\tilde{k}}^{\tilde{a}\tilde{b}\tilde{c}'} \rangle = - \bar{h}_{\tilde{c}\tilde{i}j}^{\tilde{i}\tilde{c}'\tilde{j}} - \bar{h}_{\tilde{c}\tilde{i}k}^{\tilde{i}\tilde{c}'\tilde{k}} - \bar{h}_{\tilde{c}\tilde{j}k}^{\tilde{j}\tilde{c}'\tilde{k}}$$
$$+ \bar{h}_{\tilde{a}\tilde{i}\tilde{c}}^{\tilde{i}\tilde{c}'\tilde{a}} + \bar{h}_{\tilde{c}\tilde{i}b}^{\tilde{i}\tilde{c}'\tilde{b}}$$
$$+ \bar{h}_{\tilde{a}\tilde{j}\tilde{c}}^{\tilde{j}\tilde{c}'\tilde{a}} + \bar{h}_{\tilde{c}\tilde{j}b}^{\tilde{j}\tilde{c}'\tilde{b}}$$
$$+ \bar{h}_{\tilde{a}\tilde{k}\tilde{c}}^{\tilde{k}\tilde{c}'\tilde{a}} + \bar{h}_{\tilde{c}\tilde{k}\tilde{b}}^{\tilde{k}\tilde{c}'\tilde{b}} \tag{10}$$

$$\langle \Phi_{ijk}^{abc} | \bar{H}_3 | \Phi_{ijk'}^{abc} \rangle = - \bar{h}_{aik'}^{iak} - \bar{h}_{ajk'}^{jak}$$
$$- \bar{h}_{bik'}^{ibk} - \bar{h}_{bjk'}^{jbk}$$
$$- \bar{h}_{cik'}^{ick} - \bar{h}_{cjk'}^{jck}$$
$$+ \bar{h}_{ak'b}^{kab} + \bar{h}_{ak'c}^{kac} + \bar{h}_{bk'c}^{kbc} \tag{11}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}} | \bar{H}_3 | \Phi_{ij'\tilde{k}}^{ab\tilde{c}} \rangle = - \bar{h}_{aij'}^{iaj} - \bar{h}_{aj'\tilde{k}}^{ja\tilde{k}}$$
$$- \bar{h}_{bij'}^{ibj} - \bar{h}_{bj'\tilde{k}}^{jb\tilde{k}}$$
$$- \bar{h}_{\tilde{c}j'\tilde{k}}^{j\tilde{c}\tilde{k}}$$
$$+ \bar{h}_{aj'b}^{jab} + \bar{h}_{aj'\tilde{c}}^{ja\tilde{c}} + \bar{h}_{bj'\tilde{c}}^{jb\tilde{c}} \tag{12}$$

$$\langle \Phi_{ij\tilde{k}}^{a\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{i'j\tilde{k}}^{a\tilde{b}\tilde{c}} \rangle = - \bar{h}_{ai'\tilde{j}}^{ia\tilde{j}} - \bar{h}_{ai'\tilde{k}}^{ia\tilde{k}}$$
$$- \bar{h}_{\tilde{b}i'\tilde{j}}^{i\tilde{b}\tilde{j}} - \bar{h}_{\tilde{b}i'\tilde{k}}^{i\tilde{b}\tilde{k}}$$
$$- \bar{h}_{\tilde{c}i'\tilde{j}}^{i\tilde{c}\tilde{j}} - \bar{h}_{\tilde{c}i'\tilde{k}}^{i\tilde{c}\tilde{k}}$$
$$+ \bar{h}_{ai'\tilde{b}}^{ia\tilde{b}} + \bar{h}_{ai'\tilde{c}}^{ia\tilde{c}} \tag{13}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}} | \bar{H}_3 | \Phi_{ij\tilde{k}'}^{ab\tilde{c}} \rangle = - \bar{h}_{aik'}^{ia\tilde{k}} - \bar{h}_{ajk'}^{ja\tilde{k}}$$
$$- \bar{h}_{bik'}^{ib\tilde{k}} - \bar{h}_{bjk'}^{jb\tilde{k}}$$
$$- \bar{h}_{\tilde{c}ik'}^{i\tilde{c}\tilde{k}} - \bar{h}_{\tilde{c}jk'}^{j\tilde{c}\tilde{k}}$$
$$+ \bar{h}_{ak'\tilde{c}}^{\tilde{k}a\tilde{c}} + \bar{h}_{bk'\tilde{c}}^{\tilde{k}b\tilde{c}} \tag{14}$$

$$
\langle \Phi^{a\tilde{b}\tilde{c}}_{i\tilde{j}\tilde{k}} | \bar{H}_3 | \Phi^{a\tilde{b}\tilde{c}}_{i\tilde{j}\tilde{k}'} \rangle = - \bar{h}^{j\tilde{a}\tilde{k}}_{a\tilde{j}\tilde{k}'}
$$

$$
- \bar{h}^{i\tilde{b}\tilde{k}}_{\tilde{b}i\tilde{k}'} - \bar{h}^{\tilde{j}\tilde{b}\tilde{k}}_{\tilde{b}\tilde{j}\tilde{k}'}
$$

$$
- \bar{h}^{i\tilde{c}\tilde{k}}_{\tilde{c}i\tilde{k}'} - \bar{h}^{\tilde{j}\tilde{c}\tilde{k}}_{\tilde{c}\tilde{j}\tilde{k}'} \tag{15}
$$

$$
+ \bar{h}^{\tilde{k}a\tilde{b}}_{a\tilde{k}'\tilde{b}} + \bar{h}^{\tilde{k}a\tilde{c}}_{a\tilde{k}'\tilde{c}} + \bar{h}^{\tilde{k}\tilde{b}\tilde{c}}_{\tilde{b}\tilde{k}'\tilde{c}}
$$

$$
\langle \Phi^{\tilde{a}\tilde{b}\tilde{c}}_{\tilde{i}\tilde{j}\tilde{k}} | \bar{H}_3 | \Phi^{\tilde{a}\tilde{b}\tilde{c}}_{\tilde{i}\tilde{j}\tilde{k}'} \rangle = - \bar{h}^{\tilde{i}\tilde{a}\tilde{k}}_{\tilde{a}\tilde{i}\tilde{k}'} - \bar{h}^{\tilde{j}\tilde{a}\tilde{k}}_{\tilde{a}\tilde{j}\tilde{k}'}
$$

$$
- \bar{h}^{\tilde{i}\tilde{b}\tilde{k}}_{\tilde{b}\tilde{i}\tilde{k}'} - \bar{h}^{\tilde{j}\tilde{b}\tilde{k}}_{\tilde{b}\tilde{j}\tilde{k}'}
$$

$$
- \bar{h}^{\tilde{i}\tilde{c}\tilde{k}}_{\tilde{c}\tilde{i}\tilde{k}'} - \bar{h}^{\tilde{j}\tilde{c}\tilde{k}}_{\tilde{c}\tilde{j}\tilde{k}'} \tag{16}
$$

$$
+ \bar{h}^{\tilde{k}\tilde{a}\tilde{b}}_{\tilde{a}\tilde{k}'\tilde{b}} + \bar{h}^{\tilde{k}\tilde{a}\tilde{c}}_{\tilde{a}\tilde{k}'\tilde{c}} + \bar{h}^{\tilde{k}\tilde{b}\tilde{c}}_{\tilde{b}\tilde{k}'\tilde{c}}
$$

A.2.3 Off-diagonal Elements with two Spin-Orbital Differences

$$\langle \Phi_{ijk}^{abc} | \bar{H}_3 | \Phi_{ijk}^{ab'c'} \rangle = + \bar{h}_{bic}^{ib'c'} + \bar{h}_{bjc}^{jb'c'} + \bar{h}_{bkc}^{kb'c'} \tag{17}$$

$$\langle \Phi_{ij\tilde{k}}^{abc} | \bar{H}_3 | \Phi_{ij\tilde{k}}^{a'b'\tilde{c}} \rangle = + \bar{h}_{aib}^{ia'b'} + \bar{h}_{ajb}^{ja'b'} \tag{18}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}} | \bar{H}_3 | \Phi_{ij\tilde{k}}^{ab'\tilde{c}'} \rangle = + \bar{h}_{bi\tilde{c}}^{ib'\tilde{c}'} + \bar{h}_{bj\tilde{c}}^{jb'\tilde{c}'} + \bar{h}_{b\tilde{k}\tilde{c}}^{\tilde{k}b'\tilde{c}'} \tag{19}$$

$$\langle \Phi_{ij\tilde{k}}^{a\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{ij\tilde{k}}^{a'\tilde{b}\tilde{c}'} \rangle = + \bar{h}_{ai\tilde{c}}^{ia'\tilde{c}'} + \bar{h}_{aj\tilde{c}}^{\tilde{j}a'\tilde{c}'} + \bar{h}_{ak\tilde{c}}^{\tilde{k}a'\tilde{c}'} \tag{20}$$

$$\langle \Phi_{ij\tilde{k}}^{a\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{ij\tilde{k}}^{a\tilde{b}'\tilde{c}'} \rangle = + \bar{h}_{\tilde{b}j\tilde{c}}^{\tilde{j}\tilde{b}'\tilde{c}'} + \bar{h}_{\tilde{b}\tilde{k}\tilde{c}}^{\tilde{k}\tilde{b}'\tilde{c}'} \tag{21}$$

$$\langle \Phi_{ij\tilde{k}}^{\tilde{a}\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{ij\tilde{k}}^{\tilde{a}\tilde{b}'\tilde{c}'} \rangle = + \bar{h}_{\tilde{b}i\tilde{c}}^{\tilde{i}\tilde{b}'\tilde{c}'} + \bar{h}_{\tilde{b}j\tilde{c}}^{\tilde{j}\tilde{b}'\tilde{c}'} + \bar{h}_{\tilde{b}\tilde{k}\tilde{c}}^{\tilde{k}\tilde{b}'\tilde{c}'} \tag{22}$$

$$\begin{aligned}
\langle \Phi_{ijk}^{abc} | \bar{H}_3 | \Phi_{ijk'}^{abc'} \rangle &= - \bar{h}_{cik'}^{ic'k} - \bar{h}_{cjk'}^{jc'k} \\
&\quad + \bar{h}_{ak'c}^{kac'} + \bar{h}_{bk'c}^{kbc'}
\end{aligned} \tag{23}$$

$$\begin{aligned}
\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}} | \bar{H}_3 | \Phi_{ij'\tilde{k}}^{ab'\tilde{c}} \rangle &= - \bar{h}_{bij'}^{ib'j} - \bar{h}_{bj'\tilde{k}}^{jb'\tilde{k}} \\
&\quad + \bar{h}_{aj'b}^{jab'} + \bar{h}_{bj'\tilde{c}}^{jb'\tilde{c}}
\end{aligned} \tag{24}$$

$$\begin{aligned}
\langle \Phi_{ij\tilde{k}}^{a\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{i'j\tilde{k}}^{a'\tilde{b}\tilde{c}} \rangle &= - \bar{h}_{ai'\tilde{j}}^{ia'\tilde{j}} - \bar{h}_{ai'\tilde{k}}^{ia'\tilde{k}} \\
&\quad + \bar{h}_{ai'\tilde{b}}^{ia'\tilde{b}} + \bar{h}_{ai'\tilde{c}}^{ia'\tilde{c}}
\end{aligned} \tag{25}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ijk'}^{abc'}\rangle = -\bar{h}_{\tilde{c}ik'}^{ic'\tilde{k}} - \bar{h}_{\tilde{c}jk'}^{jc'\tilde{k}}$$
$$+ \bar{h}_{ak'\tilde{c}}^{\tilde{k}ac'} + \bar{h}_{bk'\tilde{c}}^{\tilde{k}bc'} \tag{26}$$

$$\langle \Phi_{i\tilde{j}\tilde{k}}^{a\tilde{b}\tilde{c}}|\bar{H}_3|\Phi_{ij'\tilde{k}}^{ab'\tilde{c}'}\rangle = -\bar{h}_{\tilde{b}ij'}^{ib'\tilde{j}} - \bar{h}_{\tilde{b}j'\tilde{k}}^{\tilde{j}b'\tilde{k}}$$
$$+ \bar{h}_{aj'\tilde{b}}^{\tilde{j}ab'} + \bar{h}_{\tilde{b}j'\tilde{c}}^{\tilde{j}b'\tilde{c}} \tag{27}$$

$$\langle \Phi_{\tilde{i}\tilde{j}\tilde{k}}^{\tilde{a}\tilde{b}\tilde{c}}|\bar{H}_3|\Phi_{i'\tilde{j}\tilde{k}}^{a'\tilde{b}\tilde{c}'}\rangle = -\bar{h}_{\tilde{a}i'\tilde{j}}^{\tilde{i}a'\tilde{j}} - \bar{h}_{tldeai'\tilde{k}}^{\tilde{i}a'\tilde{k}}$$
$$+ \bar{h}_{\tilde{a}i'\tilde{b}}^{\tilde{i}a'\tilde{b}} + \bar{h}_{\tilde{a}i'\tilde{c}}^{\tilde{i}a'\tilde{c}} \tag{28}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ijk'}^{ab'\tilde{c}}\rangle = -\bar{h}_{bik'}^{ib'\tilde{k}} - \bar{h}_{bjk'}^{jb'\tilde{k}}$$
$$+ \bar{h}_{bk'\tilde{c}}^{\tilde{k}b'\tilde{c}} \tag{29}$$

$$\langle \Phi_{i\tilde{j}\tilde{k}}^{a\tilde{b}\tilde{c}}|\bar{H}_3|\Phi_{ij\tilde{k}'}^{a'\tilde{b}\tilde{c}}\rangle = -\bar{h}_{aik'}^{ia'\tilde{k}}$$
$$+ \bar{h}_{ak'\tilde{b}}^{\tilde{k}a'\tilde{b}} + \bar{h}_{ak'\tilde{c}}^{\tilde{k}a'\tilde{c}} \tag{30}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ij'\tilde{k}}^{abc'}\rangle = -\bar{h}_{\tilde{c}j'\tilde{k}}^{jc'\tilde{k}}$$
$$+ \bar{h}_{aj'\tilde{c}}^{jac'} + \bar{h}_{bj'\tilde{c}}^{jbc'} \tag{31}$$

$$\langle \Phi_{i\tilde{j}\tilde{k}}^{a\tilde{b}\tilde{c}}|\bar{H}_3|\Phi_{i'\tilde{j}\tilde{k}}^{a\tilde{b}\tilde{c}'}\rangle = -\bar{h}_{\tilde{c}i'\tilde{j}}^{ic'\tilde{j}} - \bar{h}_{\tilde{c}i'\tilde{k}}^{ic'\tilde{k}}$$
$$+ \bar{h}_{aic}^{iac} + \bar{h}_{\tilde{b}i'\tilde{c}}^{i\tilde{b}\tilde{c}'} \tag{32}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ij\tilde{k}'}^{abc'}\rangle = -\bar{h}_{\tilde{c}ik'}^{ic'\tilde{k}} - \bar{h}_{\tilde{c}jk'}^{jc'\tilde{k}}$$
$$+ \bar{h}_{ak'\tilde{c}}^{\tilde{k}ac'} + \bar{h}_{bk'\tilde{c}}^{\tilde{k}bc'} \tag{33}$$

$$\langle \Phi_{ij\tilde{k}}^{a\tilde{b}\tilde{c}}|\bar{H}_3|\Phi_{ij\tilde{k}'}^{a\tilde{b}\tilde{c}'}\rangle = -\,\bar{h}_{\tilde{c}i k'}^{i\tilde{c}'\tilde{k}} - \bar{h}_{\tilde{c}\tilde{j}\tilde{k}'}^{\tilde{j}\tilde{c}'\tilde{k}}$$
$$+\,\bar{h}_{a\tilde{k}'\tilde{c}}^{\tilde{k}a\tilde{c}'} + \bar{h}_{\tilde{b}\tilde{k}'\tilde{c}}^{\tilde{k}\tilde{b}\tilde{c}'} \tag{34}$$

$$\langle \Phi_{\tilde{i}\tilde{j}\tilde{k}}^{\tilde{a}\tilde{b}\tilde{c}}|\bar{H}_3|\Phi_{\tilde{i}\tilde{j}\tilde{k}'}^{\tilde{a}\tilde{b}\tilde{c}'}\rangle = -\,\bar{h}_{\tilde{c}\tilde{i}\tilde{k}'}^{\tilde{i}\tilde{c}'\tilde{k}} - \bar{h}_{\tilde{c}\tilde{j}\tilde{k}'}^{\tilde{j}\tilde{c}'\tilde{k}}$$
$$+\,\bar{h}_{\tilde{a}\tilde{k}'\tilde{c}}^{\tilde{k}\tilde{a}\tilde{c}'} + \bar{h}_{\tilde{b}\tilde{k}'\tilde{c}}^{\tilde{k}\tilde{b}\tilde{c}'} \tag{35}$$

$$\langle \Phi_{ijk}^{abc}|\bar{H}_3|\Phi_{ij\tilde{k}'}^{abc'}\rangle = -\,\bar{h}_{ci\tilde{k}'}^{ic'k} - \bar{h}_{cj\tilde{k}'}^{jc'k}$$
$$+\,\bar{h}_{a\tilde{k}'c}^{kac'} + \bar{h}_{b\tilde{k}'c}^{kbc'} \tag{36}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ij'\tilde{k}}^{ab'c'}\rangle = -\,\bar{h}_{bij'}^{ib'j} - \bar{h}_{bj'\tilde{k}}^{jb'\tilde{k}}$$
$$+\,\bar{h}_{aj'b}^{jab'} + \bar{h}_{bj'\tilde{c}}^{jb'\tilde{c}} \tag{37}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{i'\tilde{j}\tilde{k}}^{a'\tilde{b}\tilde{c}}\rangle = -\,\bar{h}_{ai'\tilde{j}}^{ia'\tilde{j}} - \bar{h}_{ak'\tilde{k}}^{ia'\tilde{k}}$$
$$+\,\bar{h}_{ai'\tilde{b}}^{ia'\tilde{b}} + \bar{h}_{ai'\tilde{c}}^{ia'\tilde{c}} \tag{38}$$

$$\langle \Phi_{ijk}^{abc}|\bar{H}_3|\Phi_{ij'k'}^{abc}\rangle = -\,\bar{h}_{aj'k'}^{jak} - \bar{h}_{bj'k'}^{jbk} - \bar{h}_{cj'k'}^{jck} \tag{39}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{i'j'\tilde{k}}^{ab\tilde{c}}\rangle = -\,\bar{h}_{ai'j'}^{iaj} - \bar{h}_{bi'j'}^{ibj} \tag{40}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ij'\tilde{k}'}^{ab\tilde{c}}\rangle = -\,\bar{h}_{aj'\tilde{k}'}^{ja\tilde{k}} - \bar{h}_{bj'\tilde{k}'}^{jb\tilde{k}} - \bar{h}_{\tilde{c}j'\tilde{k}'}^{j\tilde{c}\tilde{k}} \tag{41}$$

$$\langle \Phi_{ij\tilde{k}}^{a\tilde{b}\tilde{c}}|\bar{H}_3|\Phi_{i'\tilde{j}\tilde{k}'}^{a\tilde{b}\tilde{c}}\rangle = -\,\bar{h}_{ai'\tilde{k}'}^{ia\tilde{k}} - \bar{h}_{\tilde{b}i'\tilde{k}'}^{i\tilde{b}\tilde{k}} - \bar{h}_{\tilde{c}i'\tilde{k}'}^{i\tilde{c}\tilde{k}} \tag{42}$$

$$\langle \Phi^{a\tilde{b}\tilde{c}}_{i\tilde{j}\tilde{k}} | \bar{H}_3 | \Phi^{a\tilde{b}\tilde{c}}_{ij'\tilde{k}'} \rangle = - \bar{h}^{\tilde{i}\tilde{b}\tilde{k}}_{\tilde{b}i'\tilde{k}'} - \bar{h}^{\tilde{i}\tilde{c}\tilde{k}}_{\tilde{c}i'\tilde{k}'} \tag{43}$$

$$\langle \Phi^{\tilde{a}\tilde{b}\tilde{c}}_{i\tilde{j}\tilde{k}} | \bar{H}_3 | \Phi^{\tilde{a}\tilde{b}\tilde{c}}_{i\tilde{j}'\tilde{k}'} \rangle = - \bar{h}^{\tilde{i}\tilde{a}\tilde{k}}_{\tilde{a}i'\tilde{k}'} - \bar{h}^{\tilde{i}\tilde{b}\tilde{k}}_{\tilde{b}i'\tilde{k}'} - \bar{h}^{\tilde{i}\tilde{c}\tilde{k}}_{\tilde{c}i'\tilde{k}'} \tag{44}$$

A.2.4 Off-diagonal Elements with three Spin-Orbital Differences

$$\langle \Phi_{ijk}^{abc}|\bar{H}_3|\Phi_{ijk'}^{ab'c'}\rangle = + \bar{h}_{bk'c}^{kb'c'} \tag{45}$$

$$\langle \Phi_{ijk}^{abc}|\bar{H}_3|\Phi_{ij\tilde{k}'}^{ab'\tilde{c}'}\rangle = + \bar{h}_{b\tilde{k}''c}^{kb'\tilde{c}'} \tag{46}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ij'\tilde{k}}^{a\tilde{b}'\tilde{c}'}\rangle = + \bar{h}_{bj''\tilde{c}}^{j\tilde{b}'\tilde{c}'} \tag{47}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ij'\tilde{k}}^{ab'\tilde{c}'}\rangle = + \bar{h}_{bj'\tilde{c}}^{jb'\tilde{c}'} \tag{48}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ij'\tilde{k}}^{a'b'\tilde{c}}\rangle = + \bar{h}_{aj'b}^{ja'b'} \tag{49}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ijk'}^{ab'c'}\rangle = + \bar{h}_{bk'\tilde{c}}^{\tilde{k}b'c'} \tag{50}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ijk'}^{ab'\tilde{c}'}\rangle = + \bar{h}_{bk'\tilde{c}}^{\tilde{k}b'\tilde{c}'} \tag{51}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}}|\bar{H}_3|\Phi_{ij'\tilde{k}}^{a'b'\tilde{c}}\rangle = + \bar{h}_{aj'b}^{ja'b'} \tag{52}$$

$$\langle \Phi_{ij\tilde{k}}^{a\tilde{b}\tilde{c}}|\bar{H}_3|\Phi_{i\tilde{j}k'}^{a\tilde{b}'c'}\rangle = + \bar{h}_{\tilde{b}k'\tilde{c}}^{\tilde{k}\tilde{b}'c'} \tag{53}$$

$$\langle \Phi_{i\tilde{j}\tilde{k}}^{a\tilde{b}\tilde{c}}|\bar{H}_3|\Phi_{ij'\tilde{k}}^{ab'\tilde{c}'}\rangle = + \bar{h}_{\tilde{b}j'\tilde{c}}^{\tilde{j}b'\tilde{c}'} \tag{54}$$

$$\langle \Phi_{i\tilde{j}\tilde{k}}^{a\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{i'\tilde{j}\tilde{k}}^{\tilde{a'}\tilde{b}\tilde{c'}} \rangle = + \bar{h}_{ai'\tilde{c}}^{i\tilde{a'}\tilde{c'}} \tag{55}$$

$$\langle \Phi_{i\tilde{j}k}^{a\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{i\tilde{j}\tilde{k'}}^{a'\tilde{b}\tilde{c'}} \rangle = + \bar{h}_{ak'\tilde{c}}^{\tilde{k}a'\tilde{c'}} \tag{56}$$

$$\langle \Phi_{i\tilde{j}\tilde{k}}^{a\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{ij'\tilde{k}}^{a'b'\tilde{c}} \rangle = + \bar{h}_{aj'\tilde{b}}^{\tilde{j}a'b'} \tag{57}$$

$$\langle \Phi_{i\tilde{j}\tilde{k}}^{a\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{i'\tilde{j}\tilde{k}}^{a'\tilde{b}\tilde{c'}} \rangle = + \bar{h}_{ai'\tilde{c}}^{ia'\tilde{c'}} \tag{58}$$

$$\langle \Phi_{\tilde{i}\tilde{j}\tilde{k}}^{\tilde{a}\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{i'\tilde{j}\tilde{k}}^{a'\tilde{b}\tilde{c'}} \rangle = + \bar{h}_{\tilde{a}i'\tilde{c}}^{\tilde{i}a'\tilde{c'}} \tag{59}$$

$$\langle \Phi_{\tilde{i}\tilde{j}\tilde{k}}^{\tilde{a}\tilde{b}\tilde{c}} | \bar{H}_3 | \Phi_{\tilde{i}\tilde{j}\tilde{k'}}^{\tilde{a}\tilde{b'}\tilde{c'}} \rangle = + \bar{h}_{\tilde{b}\tilde{k'}\tilde{c}}^{\tilde{k}\tilde{b'}\tilde{c'}} \tag{60}$$

$$\langle \Phi_{ijk}^{abc} | \bar{H}_3 | \Phi_{ij'k'}^{abc'} \rangle = - \bar{h}_{cj'k'}^{jc'k} \tag{61}$$

$$\langle \Phi_{ij\tilde{k}}^{abc} | \bar{H}_3 | \Phi_{ij'\tilde{k'}}^{abc'} \rangle = - \bar{h}_{cj'\tilde{k'}}^{jc'k} \tag{62}$$

$$\langle \Phi_{ij\tilde{k}}^{ab\tilde{c}} | \bar{H}_3 | \Phi_{i'j'\tilde{k}}^{ab'\tilde{c}} \rangle = - \bar{h}_{bi'j'}^{ib'j} \tag{63}$$

$$\langle \Phi^{ab\tilde{c}}_{ij\tilde{k}} | \bar{H}_3 | \Phi^{ab'\tilde{c}}_{ij'\tilde{k}'} \rangle = - \bar{h}^{jb'\tilde{k}}_{bj'\tilde{k}'} \qquad (64)$$

$$\langle \Phi^{ab\tilde{c}}_{ij\tilde{k}} | \bar{H}_3 | \Phi^{abc'}_{ij'k'} \rangle = - \bar{h}^{jc'\tilde{k}}_{\tilde{c}j'k'} \qquad (65)$$

$$\langle \Phi^{ab\tilde{c}}_{ij\tilde{k}} | \bar{H}_3 | \Phi^{a\tilde{b}'\tilde{c}}_{i'\tilde{j}'\tilde{k}} \rangle = - \bar{h}^{i\tilde{b}'j}_{bi'\tilde{j}'} \qquad (66)$$

$$\langle \Phi^{ab\tilde{c}}_{ij\tilde{k}} | \bar{H}_3 | \Phi^{ab\tilde{c}'}_{ij'\tilde{k}'} \rangle = - \bar{h}^{j\tilde{c}'\tilde{k}}_{\tilde{c}j'\tilde{k}'} \qquad (67)$$

$$\langle \Phi^{ab\tilde{c}}_{ij\tilde{k}} | \bar{H}_3 | \Phi^{a\tilde{b}'\tilde{c}}_{ij'\tilde{k}'} \rangle = - \bar{h}^{j\tilde{b}'\tilde{k}}_{bj'\tilde{k}'} \qquad (68)$$

$$\langle \Phi^{a\tilde{b}\tilde{c}}_{ij\tilde{k}} | \bar{H}_3 | \Phi^{a'\tilde{b}\tilde{c}}_{i'\tilde{j}\tilde{k}'} \rangle = - \bar{h}^{ia'\tilde{k}}_{ai'\tilde{k}'} \qquad (69)$$

$$\langle \Phi^{a\tilde{b}\tilde{c}}_{ij\tilde{k}} | \bar{H}_3 | \Phi^{ab'\tilde{c}}_{i'j'\tilde{k}} \rangle = - \bar{h}^{ib'\tilde{j}}_{\tilde{b}i'j'} \qquad (70)$$

$$\langle \Phi^{a\tilde{b}\tilde{c}}_{ij\tilde{k}} | \bar{H}_3 | \Phi^{a\tilde{b}\tilde{c}'}_{i'\tilde{j}\tilde{k}'} \rangle = - \bar{h}^{i\tilde{c}'\tilde{k}}_{\tilde{c}i'\tilde{k}'} \qquad (71)$$

$$\langle \Phi^{a\tilde{b}\tilde{c}}_{ij\tilde{k}} | \bar{H}_3 | \Phi^{\tilde{a}'\tilde{b}\tilde{c}}_{i'\tilde{j}\tilde{k}'} \rangle = - \bar{h}^{i\tilde{a}'\tilde{k}}_{ci'\tilde{k}'} \qquad (72)$$

$$\langle\Phi^{a\tilde{b}\tilde{c}}_{i\tilde{j}\tilde{k}}|\bar{H}_3|\Phi^{ab'\tilde{c}}_{ij'\tilde{k}'}\rangle = -\,\bar{h}^{\tilde{j}b'\tilde{k}}_{\tilde{b}j'\tilde{k}'} \tag{73}$$

$$\langle\Phi^{a\tilde{b}\tilde{c}}_{i\tilde{j}\tilde{k}}|\bar{H}_3|\Phi^{a\tilde{b}\tilde{c}'}_{ij'\tilde{k}'}\rangle = -\,\bar{h}^{\tilde{j}\tilde{c}'\tilde{k}}_{\tilde{c}j'\tilde{k}'} \tag{74}$$

$$\langle\Phi^{\tilde{a}\tilde{b}\tilde{c}}_{\tilde{i}\tilde{j}\tilde{k}}|\bar{H}_3|\Phi^{a'\tilde{b}\tilde{c}}_{i'\tilde{j}\tilde{k}'}\rangle = -\,\bar{h}^{\tilde{i}a'\tilde{k}}_{\tilde{a}i'\tilde{k}'} \tag{75}$$

$$\langle\Phi^{\tilde{a}\tilde{b}\tilde{c}}_{i\tilde{j}\tilde{k}}|\bar{H}_3|\Phi^{\tilde{a}\tilde{b}\tilde{c}'}_{\tilde{i}j'\tilde{k}'}\rangle = -\,\bar{h}^{\tilde{j}\tilde{c}'\tilde{k}}_{\tilde{c}j'\tilde{k}'} \tag{76}$$

A.3 Hugenholtz and Brandow Diagrams Corresponding to the Similarity Transformed Hamiltonian with up to Doubly Excited Cluster Component

Table .8: Hugenholtz and Brandow diagrams, as well as their corresponding algebraic expressions resulting from the semi-disconnected terms of $\langle \Phi_{ij}^{ab} | [V_N, T_1 T_3] | \Phi \rangle$, where the indices, i, j, a, b, represent occupied (i, j) and unoccupied (a, b) α-spin orbitals.

$$\frac{1}{4} \mathscr{A}_{ij}^{ab} v_{mn}^{ef} t_{aef}^{imn} t_b^j \qquad \mathscr{A}_{ij}^{ab} v_{m\tilde{n}}^{e\tilde{f}} t_{ae\tilde{f}}^{im\tilde{n}} t_b^j \qquad \frac{1}{4} \mathscr{A}_{ij}^{ab} v_{\tilde{m}\tilde{n}}^{\tilde{e}\tilde{f}} t_{a\tilde{e}\tilde{f}}^{i\tilde{m}\tilde{n}} t_b^j$$

Table .9: Hugenholtz and Brandow diagrams, as well as their corresponding algebraic expressions resulting from the semi-disconnected terms of $\frac{1}{2}\langle\Phi_{ij}^{ab}|[V_N,(T_2)^2]|\Phi\rangle$, where the indices, i,j,a,b, represent occupied (i,j) and unoccupied (a,b) α-spin orbitals.

$$\frac{1}{4}v_{mn}^{ef}t_{ef}^{mn}t_{ab}^{ij} \qquad v_{m\tilde{n}}^{e\tilde{f}}t_{e\tilde{f}}^{m\tilde{n}}t_{ab}^{ij} \qquad \frac{1}{4}v_{\tilde{m}\tilde{n}}^{\tilde{e}\tilde{f}}t_{\tilde{e}\tilde{f}}^{\tilde{m}\tilde{n}}t_{ab}^{ij}$$

Table .10: Hugenholtz and Brandow diagrams, as well as their corresponding algebraic expressions resulting from the semi-disconnected terms of $\frac{1}{2}\langle\Phi_{ij}^{ab}|[V_N,(T_1)^2T_2]|\Phi\rangle$, where the indices, i,j,a,b, represent occupied (i,j) and unoccupied (a,b) α-spin orbitals.

$$\frac{1}{4}\mathscr{A}_{ij}v_{mn}^{ef}t_{ef}^{mn}t_a^it_b^j \qquad \mathscr{A}_{ij}v_{m\tilde{n}}^{e\tilde{f}}t_{e\tilde{f}}^{m\tilde{n}}t_a^it_b^j \qquad \mathscr{A}_{ij}v_{\tilde{m}\tilde{n}}^{\tilde{e}\tilde{f}}t_{\tilde{e}\tilde{f}}^{\tilde{m}\tilde{n}}t_a^it_b^j$$

$$\frac{1}{2}v^{ef}_{mn}t^m_e t^n_f t^{ij}_{ab}$$

$$v^{e\tilde{f}}_{m\tilde{n}}t^m_e t^{\tilde{n}}_{\tilde{f}} t^{ij}_{ab}$$

$$\frac{1}{2}v^{\tilde{e}\tilde{f}}_{\tilde{m}\tilde{n}}t^{\tilde{m}}_{\tilde{e}} t^{\tilde{n}}_{\tilde{f}} t^{ij}_{ab}$$

$$-\frac{1}{2}\mathscr{A}_{ij}^{ab}v_{mn}^{ef}t_{af}^{mn}t_{e}^{i}t_{b}^{j} \qquad -\mathscr{A}_{ij}^{ab}v_{m\tilde{n}}^{e\tilde{f}}t_{a\tilde{f}}^{m\tilde{n}}t_{e}^{i}t_{b}^{j} \qquad -\frac{1}{2}\mathscr{A}^{ab}v_{mn}^{ef}t_{ef}^{in}t_{a}^{m}t_{b}^{j} \qquad -\mathscr{A}_{ij}^{ab}v_{m\tilde{n}}^{e\tilde{f}}t_{e\tilde{f}}^{i\tilde{n}}t_{a}^{m}t_{b}^{j}$$

$$\mathscr{A}_{ij}^{ab}v_{mn}^{ef}t_{af}^{in}t_{m}^{e}t_{b}^{j} \qquad \mathscr{A}_{ij}^{ab}v_{\tilde{m}n}^{\tilde{e}f}t_{af}^{in}t_{\tilde{m}}^{\tilde{e}}t_{b}^{j} \qquad \mathscr{A}_{ij}^{ab}v_{m\tilde{n}}^{e\tilde{f}}t_{a\tilde{f}}^{i\tilde{n}}t_{m}^{e}t_{b}^{j} \qquad \mathscr{A}_{ij}^{ab}v_{\tilde{m}\tilde{n}}^{\tilde{e}\tilde{f}}t_{\tilde{a}\tilde{f}}^{i\tilde{n}}t_{\tilde{m}}^{\tilde{n}}t_{b}^{j}$$

Table .11: Hugenholtz and Brandow diagrams, as well as their corresponding algebraic expressions resulting from the semi-disconnected terms of $\frac{1}{24}\langle\Phi_{ij}^{ab}|[V_N,(T_1)^4]|\Phi\rangle$, where the indices, i,j,a,b, represent occupied (i,j) and unoccupied (a,b) α-spin orbitals.

$$\mathcal{A}_{ij}v_{mn}^{ef}t_e^m t_f^n t_a^i t_b^j \quad \mathcal{A}_{ij}v_{m\tilde{n}}^{e\tilde{f}}t_e^m t_{\tilde{f}}^{\tilde{n}}t_a^i t_b^j \quad \mathcal{A}_{ij}v_{\tilde{m}\tilde{n}}^{\tilde{e}\tilde{f}}t_{\tilde{e}}^{\tilde{m}}t_{\tilde{f}}^{\tilde{n}}t_a^i t_b^j \quad -\mathcal{A}_{ij}^{ab}v_{mn}^{ef}t_a^m t_e^i t_f^n t_b^j \quad -\mathcal{A}_{ij}^{ab}v_{m\tilde{n}}^{e\tilde{f}}t_a^m t_e^i t_{\tilde{f}}^{\tilde{n}}t_b^j$$

157

REFERENCES

REFERENCES

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

[2]   E. Schrödinger, *Ann. Phys.* **384**, 361 (1926).

[3]   E. Schrödinger, *Ann. Phys.* **384**, 489 (1926).

[4]   E. Schrödinger, *Ann. Phys.* **384**, 734 (1926).

[5]   E. Schrödinger, *Ann. Phys.* **385**, 437 (1926).

[6]   E. Schrödinger, *Ann. Phys.* **386**, 109 (1926).

[7]   E. Schrödinger, *Phys. Rev.* **28**, 1049 (1926).

[8]   P. A. M. Dirac, *Proc. R. Soc. Lond. A* **117**, 610 (1928).

[9]   P. A. M. Dirac, *Proc. R. Soc. Lond. A* **118**, 351 (1928).

[10] P. A. M. Dirac, *Proc. R. Soc. Lond. A* **123**, 714 (1929).

[11] P. A. M. Dirac, *Proc. R. Soc. Lond. A* **126**, 360 (1930).

[12] Ø. Burrau, *Naturwissenshaften* **15**, 16 (1927).

[13] S. R. White, *Phys. Rev. Lett.* **69**, 2863 (1992).

[14] S. R. White and R. L. Martin, *J. Chem. Phys.* **110**, 4127 (1999).

[15] A. O. Mitrushenkov, G. Fano, F. Ortolani, R. Linguerri, and P. Palmieri, *J. Chem. Phys.* **115**, 6815 (2001).

[16] G. K.-L. Chan and M. Head-Gordon, *J. Chem. Phys.* **116**, 4462 (2002).

[17] S. Keller, M. Dolfi, M. Troyer, and M. Reiher, *J. Chem. Phys.* **143**, 244118 (2015).

[18] G. K.-L. Chan, A. Keselman, N. Nakatani, Z. Li, and S. R. White, *J. Chem. Phys.* **145**, 014102 (2016).

[19] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).

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

[21] K. Burke, *J. Chem. Phys.* **136**, 150901 (2012).

[22] A. Becke, *J. Chem. Phys.* **140**, 18A301 (2014).

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

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

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

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

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

[28] J. P. Perdew and A. Ruzsinszky, *Int. J. Quantum Chem.* **110**, 2801 (2010).

[29] Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.* **120**, 215 (2008).

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

[31] J. Goldstone, *Proceddings R. Soc. London. Ser. A, Math. Phys. Sci.* **239**, 267 (1957).

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

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

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

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

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

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

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

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

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

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

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

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

[44] M. Takahashi and J. Paldus, *Phys. Rev. B* **31**, 5121 (1985).

[45] P. Piecuch, S. Zarrabian, J. Paldus, and J. Čížek, *Phys. Rev. B* **42**, 3351 (1990).

[46] P. Piecuch and J. Paldus, *Int. J. Quantum Chem. Symp.* **25**, 135 (1991).

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

[48] P. Piecuch, J. Čížek, and J. Paldus, *Int. J. Quantum Chem.* **42**, 165 (1992).

[49] R. Podeszwa, S. A. Kucharski, and L. Z. Stolarczyk, *J. Chem. Phys.* **116**, 480 (2002).

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

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

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

[53] B. Jeziorski and H. J. Monkhorst, *Phys. Rev. A* **24**, 1668 (1981).

[54] B. Jeziorski and J. Paldus, *J. Chem. Phys.* **88**, 5673 (1988).

[55] L. Meissner, K. Jankowski, and J. Wasilewski, *Int. J. Quantum Chem.* **34**, 535 (1988).

[56] P. Piecuch and J. Paldus, *Theor. Chim. Acta* **83**, 69 (1992).

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

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

[59] P. Piecuch and J. Paldus, *J. Chem. Phys.* **101**, 5875 (1994).

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

[61] K. Kowalski and P. Piecuch, *Chem. Phys. Lett.* **334**, 89 (2001).

[62] K. Kowalski and P. Piecuch, *J. Molec. Struct.: THEOCHEM* **547**, 191 (2001).

[63] K. Kowalski and P. Piecuch, *Mol. Phys.* **102**, 2425 (2004).

[64] L. Meissner, S. A. Kucharski, and R. J. Bartlett, *J. Chem. Phys.* **91**, 6187 (1989).

[65] L. Meissner and R. J. Bartlett, *J. Chem. Phys.* **92**, 561 (1990).

[66] S. A. Kucharski and R. J. Bartlett, *J. Chem. Phys.* **95**, 8227 (1991).

[67] A. Balková, S. A. Kucharski, L. Meissner, and R. J. Bartlett, *Theor. Chim. Acta* **80**, 335 (1991).

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

[69] J. Paldus and X. Li, *J. Chem. Phys.* **118**, 6769 (2003).

[70] X. Li and J. Paldus, *J. Chem. Phys.* **119**, 5320 (2003).

[71] X. Li and J. Paldus, *J. Chem. Phys.* **119**, 5334 (2003).

[72] X. Li and J. Paldus, *J. Chem. Phys.* **119**, 5346 (2003).

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

[74] P. Piecuch, X. Li, and J. Paldus, *Chem. Phys. Lett.* **230**, 377 (1994).

[75] P. Piecuch and J. Paldus, *J. Phys. Chem.* **99**, 15354 (1995).

[76] J. Paldus, P. Piecuch, B. Jeziorski, and L. Pylypow, in *Recent Progress in Many-Body Theories*, Vol. 3, edited by T. L. Ainsworthy, C. E. Campbell, B. E. Clements, and E. Krotschek (Plenum Press, New York, 1992) pp. 287–303.

[77] U. S. Mahapatra, B. Datta, and D. Mukherjee, *J. Chem. Phys.* **110**, 6171 (1999).

[78] U. S. Mahapatra, B. Datta, and D. Mukherjee, *Chem. Phys. Lett.* **299**, 42 (1999).

[79] J. Masik and I. Hubač, *Adv. Quantum Chem.* **31**, 75 (1998).

[80] J. Pittner, P. Nachtigall, P. Čarsky, J. Masik, and I. Hubač, *J. Chem. Phys.* **110**, 10275 (1999).

[81] I. Hubač, J. Pittner, and P. Čarsky, *J. Chem. Phys.* **112**, 8779 (2000).

[82] J. Pittner, *J. Chem. Phys.* **118**, 10876 (2003).

[83] M. Hanrath, *J. Chem. Phys.* **123**, 084102 (2005).

[84] M. Hanrath, *Chem. Phys. Lett.* **420**, 426 (2006).

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

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

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

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

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

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

[91] K. Kowalski, S. Hirata, M. Włoch, P. Piecuch, and T. L. Windus, *J. Chem. Phys.* **123**, 074319 (2005).

[92] P. Piecuch, S. Hirata, K. Kowalski, P.-D. Fan, and T. L. Windus, *Int. J. Quantum Chem.* **106**, 79 (2006).

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

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

[95] K. W. Sattelmeyer, H. F. Schaefer, III, and J. F. Stanton, *Chem. Phys. Lett.* **378**, 42 (2003).

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

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

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

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

[100] P.-D. Fan, M. Kamiya, and S. Hirata, *J. Chem. Theor. Comput.* **3**, 1036 (2007).

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

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

[103] A. O. Ajala, J. Shen, and P. Piecuch, *J. Phys. Chem. A* **121**, 3469 (2017), pMID: 28409643.

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

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

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

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

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

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

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

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

[112] M. Musiał, S. A. Kucharski, and R. J. Bartlett, *J. Chem. Phys.* **116**, 4382 (2002).

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

[114] G. K.-L. Chan, M. Kállay, and J. Gauss, *J. Chem. Phys.* **121**, 6110 (2004).

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

[116] J. W. Krogh and J. Olsen, *Chem. Phys. Lett.* **344**, 578 (2001).

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

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

[119] Y. S. Lee and R. J. Bartlett, *J. Chem. Phys.* **80**, 4371 (1984).

[120] Y. S. Lee, S. A. Kucharski, and R. J. Bartlett, *J. Chem. Phys.* **81**, 5906 (1984).

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

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

[123] G. W. Trucks, J. Noga, and R. J. Bartlett, *Chem. Phys. Lett.* **145**, 548 (1988).

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

[125] M. Urban, J. Noga, S. J. Cole, and R. J. Bartlett, *J. Chem. Phys.* **83**, 4041 (1985).

[126] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, *Chem. Phys. Lett.* **102**, 479 (1989).

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

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

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

[130] O. Christiansen, H. Koch, P. Jørgensen, and J. Olsen, *Chem. Phys. Lett.* **256**, 185 (1996).

[131] H. Koch, O. Christiansen, P. Jørgensen, and J. Olsen, *Chem. Phys. Lett.* **244**, 75 (1995).

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

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

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

[135] 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.

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

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

[138] P. Piecuch, S. A. Kucharski, and K. Kowalski, *Chem. Phys. Lett.* **344**, 176 (2001).

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

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

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

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

[143] M. D. Lodriguito, K. Kowalski, M. Włoch, and P. Piecuch, *J. Mol. Struct: THEOCHEM* **771**, 89 (2006).

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

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

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

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

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

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

[150] 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).

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

[152] G. Fradelos, J. J. Lutz, T. A. Wesołowski, P. Piecuch, and M. Włoch, in *Advances in the Theory of Quantum Systems in Chemistry and Physics*, Progress in Theoretical Chemistry and Physics, Vol. 22, edited by P. E. Hoggan, E. J. Brändas, J. Maruani, P. Piecuch, and G. Delgado-Barrio (Springer, Dordrecht, 2012) pp. 219–248.

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

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

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

[156] V. Alexandrov, P. Piecuch, and L. Adamowicz, *J. Chem. Phys.* **102**, 3301 (1995).

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

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

[159] K. B. Ghose, P. Piecuch, S. Pal, and L. Adamowicz, *J. Chem. Phys.* **104**, 6582 (1996).

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

[161] P. Piecuch, S. A. Kucharski, and V. Špirko, *J. Chem. Phys.* **111**, 6679 (1999).

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

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

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

[165] 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.

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

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

[168] N. Oliphant and L. Adamowicz, *Int. Rev. Phys. Chem.* **12**, 339 (1993).

[169] P. Piecuch and L. Adamowicz, *Chem. Phys. Lett.* **221**, 121 (1993).

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

[171] K. Duanmu, O. Roberto-Neto, F. B. C. Machado, J. A. Hansen, J. Shen, P. Piecuch, and D. G. Truhlar, *J. Phys. Chem. C* **120**, 13275 (2016).

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

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

[174] N. Metropolis and S. Ulam, *J. Am. Stat. Assoc.* **44**, 335 (1949).

[175] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).

[176] W. K. Hastings, *Biometrika* **57**, 97 (1970).

[177] W. L. McMillan, *Phys. Rev.* **138**, A442 (1965).

[178] D. Schiff and L. Verlet, *Phys. Rev.* **160**, 208 (1967).

[179] D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).

[180] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, *J. Chem. Phys.* **77**, 5593 (1982).

[181] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).

[182] D. Ceperley, G. V. Chester, and M. H. Kalos, *Phys. Rev. B* **16**, 3081 (1977).

[183] D. M. Ceperley, *Rev. Mineral Geochem.* **71**, 129 (2010).

[184] J. Toulouse, R. Assaraf, and C. J. Umrigar, *Adv. Quantum Chem.* **73**, 285 (2016).

[185] J. B. Anderson, *J. Chem. Phys.* **63**, 1499 (1975).

[186] D. J. Klein and H. M. Pickett, *J. Chem. Phys.* **64**, 4811 (1976).

[187] J. B. Anderson, *J. Chem. Phys.* **65**, 1421 (1976).

[188] J. B. Anderson, *Int. J. Quantum Chem.* **15**, 109 (1979).

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

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

[191] N. S. Blunt, S. D. Smart, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **143**, 134117 (2015).

[192] G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, *Nature* **493**, 365 (2013).

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

[194] H. Luo and A. Alavi, *J. Chem. Theory Comput.* **14**, 1403 (2018).

[195] A. J. Cohen, H. Luo, K. Guther, W. Dobrautz, D. P. Tew, and A. Alavi, *J. Chem. Phys.* **151** (2019).

[196] W. Dobrautz, H. Luo, and A. Alavi, *Phys. Rev. B* **99**, 075119 (2019).

[197] A. J. Cohen, , H. Luo, K. Gunther, W. Dobrautz, D. P. Tew, and A. Alavi, *J. Chem. Phys.* **151**, 061101 (2019).

[198] A. J. W. Thom, *Phys. Rev. Lett.* **105**, 1 (2010).

[199] J. S. Spencer and A. J. W. Thom, *J. Chem. Theory Comput.* **144**, 084108 (2016).

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

[201] M.-A. Filip, C. J. C. Scott, and A. J. W. Thom, *J. Chem. Theory Comput.* **15**, 6625 (2019).

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

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

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

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

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

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

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

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

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

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

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

[213] E. Xu and S. Li, *J. Chem. Phys.* **142**, 094119 (2015).

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

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

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

[217] N. P. Bauman, *Merging active-space and renormalized coupled-cluster methods via the CC(P;Q) formalism, with applications to chemical reaction profiles and singlet–triplet gaps*, Ph.D. thesis, Michigan State University (2016).

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

[219] G. H. Booth, D. Cleland, A. J. W. Thom, and A. Alavi, *J. Chem. Phys.* **135** (2011).

[220] J. J. Shepherd, A. Grüneis, G. H. Booth, G. Kresse, and A. Alavi, *Phys. Rev. B* **86**, 1 (2012).

[221] J. J. Shepherd, G. Booth, A. Grüneis, and A. Alavi, *Phys. Rev. B* **85**, 1 (2012).

[222] J. J. Shepherd, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **136**, 244101 (2012).

[223] C. Daday, S. Smart, G. H. Booth, A. Alavi, and C. Filippi, *J. Chem. Theory Comput.* **8**, 4441 (2012).

[224] D. Cleland, G. H. Booth, C. Overy, and A. Alavi, *J. Chem. Theory Comput.* **8**, 4138 (2012).

[225] G. H. Booth, D. Cleland, A. Alavi, and D. P. Tew, *J. Chem. Phys.* **137** (2012).

[226] G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, *Nature* **493**, 365 (2013).

[227] N. S. Blunt, S. D. Smart, J. A. F. Kersten, J. S. Spencer, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **142**, 184107 (2015).

[228] H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.* **91**, 461 (1989).

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

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

[231] P. Saxe, H. F. Schaefer, III, and N. C. Handy, *Chem. Phys. Lett.* **79**, 202 (1981).

[232] R. J. Harrison and N. C. Handy, *Chem. Phys. Lett.* **95**, 386 (1983).

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

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

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

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

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

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

[239] 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. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, *J. Comput. Chem.* **14**, 1347 (1993).

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

[241] M. S. Gordon and M. W. Schmidt, in *Theory and Applications of Computational Chemistry: The First Forty Years*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Elsevier, Amsterdam, 2005) pp. 1167–1190.

[242] J. Olsen, P. Jorgensen, H. Koch, A. Balkova, and R. J. Bartlett, *J. Chem. Phys.* **104**, 8007 (1996).

[243] J. S. Spencer, N. S. Blunt, S. Choi, J. Etrych, M.-A. Filip, W. M. C. Foulkes, R. S. T. Franklin, W. J. Handley, F. D. Malone, V. A. Neufeld, R. Di Remigio, T. W. Rogers, C. J. C. Scott, J. J. Shepherd, W. A. Vigor, J. Weston, R. Xu, and A. J. W. Thom, *J. Chem. Theory Comput.* **15**, 1728 (2019).

[244] J. S. Spencer, N. S. Blunt, W. A. Vigor, F. D. Malone, W. M. C. Foulkes, J. S. Shepherd, and A. J. W. Thom, *J. Open Res. Softw.* **3**, 1 (2015).

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

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

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

[248] S. Hirata, P.-D. Fan, A. A. Auer, M. Nooijen, and P. Piecuch, *J. Chem. Phys.* **121**, 12197 (2004).

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

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

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

[252] 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.

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

[254] T. Shiozaki, K. Hirao, and S. Hirata, *J. Chem. Phys.* **126**, 224106 (2007).

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

[256] P. Piecuch, *Algebraic and Diagrammatic Methods for Many-Fermion Systems* (2016 (accessed April, 2020)).

[257] J. Paldus, *Nijmegen Lectures* (1981 (accessed April, 2020)).

[258] J. Paldus and J. Čížek, *Adv. Quantum Chem.* **9**, 105 (1975).

[259] J. Paldus, in *Methods in Computational Molecular Physics, NATO Advanced Study Institute, Series B: Physics*, Vol. 293, edited by S. Wilson and G. H. F. Diercksen (Plenum, New York, 1992) p. 99.

[260] J. Paldus, in *Encyclopedia of Computational Chemistry*, Vol. 1, edited by S. Wilson (Wiley, Chichester, 2003) p. 615.

[261] J. Paldus, *J. Chem. Phys.* **67**, 303 (1977).

[262] S. Kucharski and R. Bartlett, *Adv. Quantum Chem.* **18**, 281 (1986).

[263] M. Lodriguito and P. Piecuch, in *Frontiers in Quantum Systems in Chemistry and Physics*, Progress in Theoretical Chemistry and Physics, Vol. 18, edited by S. Wilson, P. Grout, J. Maruani, G. Delgado-Barrio, and P. Piecuch (Springer, Dordrecht, 2008) pp. 67–174.

[264] F. G. Van Zee and R. A. van de Geijn, *ACM Trans. Math. Softw.* **41**, 14:1 (2015).

[265] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 1st ed. (Dover Publications, Inc., Mineola, 1996).