

# High-level coupled-cluster energetics by Monte Carlo sampling and moment expansions: Further details and comparisons

Cite as: J. Chem. Phys. **154**, 124103 (2021); <https://doi.org/10.1063/5.0045468>

Submitted: 26 January 2021 . Accepted: 04 March 2021 . Published Online: 22 March 2021

 J. Emiliano Deustua,  Jun Shen, and  Piotr Piecuch

## COLLECTIONS

Paper published as part of the special topic on [Frontiers of Stochastic Electronic Structure Calculations](#)



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

[Recent developments in the general atomic and molecular electronic structure system](#)

The Journal of Chemical Physics **152**, 154102 (2020); <https://doi.org/10.1063/5.0005188>

[Improving molecular force fields across configurational space by combining supervised and unsupervised machine learning](#)

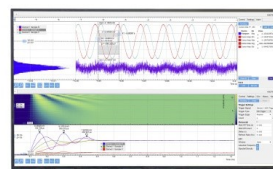
The Journal of Chemical Physics **154**, 124102 (2021); <https://doi.org/10.1063/5.0035530>

[Reliable transition properties from excited-state mean-field calculations](#)

The Journal of Chemical Physics **154**, 124106 (2021); <https://doi.org/10.1063/5.0041233>

Challenge us.

What are your needs for  
periodic signal detection?



Zurich  
Instruments

# High-level coupled-cluster energetics by Monte Carlo sampling and moment expansions: Further details and comparisons

Cite as: J. Chem. Phys. 154, 124103 (2021); doi: 10.1063/5.0045468

Submitted: 26 January 2021 • Accepted: 4 March 2021 •

Published Online: 22 March 2021



View Online



Export Citation



CrossMark

J. Emiliano Deustua,<sup>1</sup>  Jun Shen,<sup>1</sup>  and Piotr Piecuch<sup>1,2,a)</sup> 

## AFFILIATIONS

<sup>1</sup>Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, USA

<sup>2</sup>Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA

**Note:** This paper is part of the JCP Special Topic on Frontiers of Stochastic Electronic Structure Calculations.

**a)** Author to whom correspondence should be addressed: [piecuch@chemistry.msu.edu](mailto:piecuch@chemistry.msu.edu)

## ABSTRACT

We recently proposed a novel approach to converging electronic energies equivalent to high-level coupled-cluster (CC) computations by combining the deterministic  $CC(P;Q)$  formalism with the stochastic configuration interaction (CI) and CC Quantum Monte Carlo (QMC) propagations. This article extends our initial study [J. E. Deustua, J. Shen, and P. Piecuch, Phys. Rev. Lett. **119**, 223003 (2017)], which focused on recovering the energies obtained with the CC method with singles, doubles, and triples (CCSDT) using the information extracted from full CI QMC and CCSDT-MC, to the CIQMC approaches truncated at triples and quadruples. It also reports our first semi-stochastic  $CC(P;Q)$  calculations aimed at converging the energies that correspond to the CC method with singles, doubles, triples, and quadruples (CCSDTQ). The ability of the semi-stochastic  $CC(P;Q)$  formalism to recover the CCSDT and CCSDTQ energies, even when electronic quasi-degeneracies and triply and quadruply excited clusters become substantial, is illustrated by a few numerical examples, including the F–F bond breaking in  $F_2$ , the automerization of cyclobutadiene, and the double dissociation of the water molecule.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0045468>

## I. INTRODUCTION

One of the main goals of quantum chemistry is to provide an accurate and systematically improvable description of many-electron correlation effects needed to determine molecular potential energy and property surfaces and understand chemical reactivity and various types of spectroscopy. In searching for the best solutions in this area, the size extensive methods based on the exponential wave function ansatz<sup>1,2</sup> of coupled-cluster (CC) theory,<sup>3–7</sup>

$$|\Psi\rangle = e^T |\Phi\rangle, \quad (1)$$

where

$$T = \sum_{n=1}^N T_n \quad (2)$$

is the cluster operator,  $T_n$  is the  $n$ -body component of  $T$ ,  $N$  is the number of correlated electrons, and  $|\Phi\rangle$  is the reference

determinant, and their extensions to excited, open-shell, and multi-reference states<sup>8–12</sup> are among the top contenders. In this study, we focus on the higher-ranking members of the single-reference CC hierarchy beyond the basic CC singles and doubles (CCSD) level, where  $T$  is truncated at  $T_2$ ,<sup>13–16</sup> especially on the CC approach with singles, doubles, and triples (CCSDT), where  $T$  is truncated at  $T_3$ ,<sup>17–19</sup> and the CC approach with singles, doubles, triples, and quadruples (CCSDTQ), where  $T$  is truncated at  $T_4$ .<sup>20–22</sup> This is motivated by the fact that in a great many cases relevant to chemistry, including molecular properties at equilibrium geometries, multi-reference situations involving smaller numbers of strongly correlated electrons, as in the case of bond breaking and formation in the course of chemical reactions, noncovalent interactions, and photochemistry, the single-reference CCSD, CCSDT, CCSDTQ, etc. methods and their equation-of-motion (EOM)<sup>23–30</sup> and linear response<sup>31–39</sup> extensions rapidly converge to the exact, full configuration interaction (FCI) limit, allowing one to incorporate the relevant many-electron correlation effects in a conceptually

straightforward manner through particle–hole excitations from a single Slater determinant defining the Fermi vacuum without loss of accuracy as the system becomes larger characterizing truncated CI methods.<sup>10</sup>

The convergence of the single-reference CCSD, CCSDT, CCSDTQ, etc. hierarchy toward FCI in situations other than larger numbers of strongly entangled electrons is fast, but costs of the post-CCSD computations needed to achieve a quantitative description, which are determined by the iterative  $n_o^3 n_u^5$  steps in the CCSDT case and the iterative  $n_o^4 n_u^6$  steps in the case of CCSDTQ, where  $n_o$  ( $n_u$ ) is the number of occupied (unoccupied) correlated orbitals, are usually prohibitively expensive. This is why part of the CC method development effort has been devoted to finding approximate ways of incorporating higher-than-two-body components of the cluster operator  $T$ , i.e.,  $T_n$  components with  $n > 2$ , and the analogous higher-order components of the EOMCC excitation, electron-attachment, and electron-detachment operators, which could reduce enormous computational costs of the CCSDT, CCSDTQ, and similar schemes, while eliminating failures of the CCSD[T],<sup>40</sup> CCSD(T),<sup>41</sup> CCSDT-1,<sup>42,43</sup> CC3,<sup>44,45</sup> and other perturbative CC approaches (see Ref. 10 for a review) that fail when bond breaking, biradicals, and other typical multi-reference situations in chemistry are examined.<sup>8,10,46–48</sup> In fact, the analogous effort has been taking place in other areas of many-body theory, such as studies of nuclear matter, where a systematic, computationally efficient, and robust incorporation of higher-order many-particle correlation effects is every bit as important as in the case of electronic structure theory and where the quantum-chemistry-inspired CC and EOMCC methods, thanks, in part, to our group's involvement,<sup>49–55</sup> have become quite popular (see, e.g., Ref. 56 and references therein). While substantial progress in the above area, reviewed, for example, in Refs. 10, 48, and 57, has already been made, the search for the optimum solution that would allow us to obtain the results of the full CCSDT, full CCSDTQ, or similar quality at the fraction of the cost and without having to rely on perturbative concepts or user- and system-dependent ideas, such as the idea of active orbitals to select higher-than-two-body components of the cluster and EOMCC excitation operators,<sup>48</sup> continues.

In order to address this situation, we have started exploring a radically new way of converging accurate electronic energetics equivalent to those obtained with the high-level CC approaches of the full CCSDT, full CCSDTQ, and similar types, at the small fraction of the computational cost and preserving the black-box character of conventional single-reference methods, even when higher-than-two-body components of the cluster and excitation operators characterizing potential energy surfaces along bond stretching coordinates become large.<sup>58</sup> The key idea of the approach suggested in Ref. 58, which we have recently extended to excited states,<sup>59,60</sup> is a merger of the deterministic formalism, abbreviated as CC( $P$ ;Q),<sup>57,61–63</sup> which enables one to correct energies obtained with conventional as well as unconventional truncations in the cluster and EOMCC excitation operators for any category of many-electron correlation effects of interest, with the stochastic FCI Quantum Monte Carlo (FCIQMC)<sup>64–67</sup> and CC Monte Carlo (CCMC)<sup>68–71</sup> methods (see Refs. 72–74 for alternative ways of combining FCIQMC with the deterministic CC framework). As shown in Refs. 58 and 60, where we reported preliminary calculations aimed at recovering full CCSDT and EOMCCSDT<sup>26–28</sup> energetics, the

resulting semi-stochastic CC( $P$ ;Q) methodology, using the FCIQMC and CCSDT-MC approaches to identify the leading determinants or cluster amplitudes in the wave function and the *a posteriori* CC( $P$ ;Q) corrections to capture the remaining correlations, rapidly converges to the target energetics based on the information extracted from the early stages of FCIQMC or CCSDT-MC propagations. If confirmed through additional tests and comparisons involving various QMC and CC levels, the merger of the deterministic CC( $P$ ;Q) and stochastic CIQMC and CCMC ideas, originally proposed in Ref. 58, may substantially impact accurate quantum calculations for many-electron and other many-fermion systems, opening interesting new possibilities in this area.

The present study is our next step in the development and examination of the semi-stochastic CC( $P$ ;Q) methodology. In this work, we extend our initial study,<sup>58</sup> which focused on recovering the full CCSDT energetics based on the information extracted from the FCIQMC and CCSDT-MC propagations, to the CIQMC methods truncated at triples (CISDT-MC) or triples and quadruples (CISDTQ-MC), which may offer significant savings in the computational effort compared to FCIQMC and which are formally compatible with the CCSDT and CCSDTQ excitation manifolds we would like to capture. We also report our initial results of the semi-stochastic CC( $P$ ;Q) calculations aimed at converging the full CCSDTQ energetics. The ability of the semi-stochastic CC( $P$ ;Q) approaches to recover the CCSDT and CCSDTQ energies based on the truncated CISDT-MC and CISDTQ-MC propagations, even when electronic quasi-degeneracies and  $T_3$  and  $T_4$  clusters become substantial, is illustrated using the challenging cases of the F–F bond breaking in  $F_2$ , the automerization of cyclobutadiene, and the double dissociation of the water molecule as examples.

## II. THEORY AND ALGORITHMIC DETAILS

As pointed out in the Introduction, the semi-stochastic CC( $P$ ;Q) approach proposed in Ref. 58 is based on combining the deterministic CC( $P$ ;Q) framework, developed mainly in Refs. 57, 61, and 63, with the CIQMC and CCMC ideas that were originally laid down in Refs. 64, 65, and 68. Thus, we divide this section into two subsections. In Sec. II A, we summarize the key elements of the deterministic CC( $P$ ;Q) formalism, focusing on the ground-state problem relevant to the calculations reported in this study. Section II B provides information about the semi-stochastic CC( $P$ ;Q) methods developed and tested in this work, which aim at converging the CCSDT and CCSDTQ energies with the help of the FCIQMC, CISDT-MC, and CISDTQ-MC propagations.

### A. Basic elements of the ground-state CC( $P$ ;Q) formalism

The CC( $P$ ;Q) formalism has emerged out of our interest in generalizing the biorthogonal moment energy expansions, which in the past resulted in the completely renormalized (CR) CC and EOMCC approaches, including CR-CC(2,3),<sup>75–79</sup> CR-EOMCC(2,3),<sup>77,80</sup>  $\delta$ -CR-EOMCC(2,3),<sup>81</sup> and their higher-order extensions,<sup>52,63,82–84</sup> such that one can correct the CC/EOMCC energies obtained with unconventional truncations in the cluster and EOMCC excitation

operators, in addition to the conventional ones at a given many-body rank, for essentially any category of many-electron correlation effects of interest. The CC( $P$ ;Q) framework is general, i.e., it applies to ground as well as excited states, but since this work deals with the calculations that aim at recovering the ground-state CCSDT and CCSDTQ energetics, in the description below, we focus on the ground-state CC( $P$ ;Q) theory.

According to the formal CC( $P$ ;Q) prescription, the ground-state energy of a  $N$ -electron system is determined in two steps. In the initial, iterative, CC( $P$ ) step, we solve the CC equations in the subspace  $\mathcal{H}^{(P)}$  of the  $N$ -electron Hilbert space  $\mathcal{H}$ . We assume that the subspace  $\mathcal{H}^{(P)}$ , which we also call the  $P$  space, is spanned by the excited determinants  $|\Phi_K\rangle = E_K|\Phi\rangle$  that together with the reference determinant  $|\Phi\rangle$  provide the leading contributions to the target ground state  $|\Psi\rangle$  ( $E_K$  designates the usual elementary particle-hole excitation operator generating  $|\Phi_K\rangle$  from  $|\Phi\rangle$ ). In other words, we approximate the cluster operator  $T$  in Eq. (1) by

$$T^{(P)} = \sum_{|\Phi_K\rangle \in \mathcal{H}^{(P)}} t_K E_K \quad (3)$$

and solve the usual system of CC equations,

$$\mathfrak{M}_K(P) = 0, \quad |\Phi_K\rangle \in \mathcal{H}^{(P)}, \quad (4)$$

where

$$\mathfrak{M}_K(P) = \langle \Phi_K | \bar{H}^{(P)} | \Phi \rangle \quad (5)$$

are the generalized moments of the  $P$ -space CC equations<sup>85–87</sup> and

$$\bar{H}^{(P)} = e^{-T^{(P)}} H e^{T^{(P)}} = (H e^{T^{(P)}})_C \quad (6)$$

is the relevant similarity-transformed Hamiltonian, for the cluster amplitudes  $t_K$  [subscript  $C$  in Eq. (6) designates the connected operator product]. Once the cluster operator  $T^{(P)}$  and the ground-state energy

$$E^{(P)} = \langle \Phi | \bar{H}^{(P)} | \Phi \rangle \quad (7)$$

that corresponds to it are determined, we proceed to the second step of CC( $P$ ;Q) considerations, which is the calculation of the non-iterative correction  $\delta(P$ ;Q) to the CC( $P$ ) energy  $E^{(P)}$  that accounts for the many-electron correlation effects captured by another subspace of the  $N$ -electron Hilbert space  $\mathcal{H}$ , designated as  $\mathcal{H}^{(Q)}$  and called the  $Q$  space, which satisfies the condition  $\mathcal{H}^{(Q)} \subseteq (\mathcal{H}^{(0)} \oplus \mathcal{H}^{(P)})^\perp$ , where  $\mathcal{H}^{(0)}$  is a one-dimensional subspace of  $\mathcal{H}$  spanned by the reference determinant  $|\Phi\rangle$ . The formula for the  $\delta(P$ ;Q) correction is<sup>57,58,60,61,63</sup>

$$\delta(P;Q) = \sum_{\substack{|\Phi_K\rangle \in \mathcal{H}^{(Q)} \\ \text{rank}(|\Phi_K\rangle) \leq \min(N_0^{(P)}, \Xi^{(Q)})}} \ell_K(P) \mathfrak{M}_K(P), \quad (8)$$

where integer  $N^{(P)}$  defines the highest many-body rank of the excited determinants  $|\Phi_K\rangle$  relative to  $|\Phi\rangle$  [ $\text{rank}(|\Phi_K\rangle)$ ] for which

moments  $\mathfrak{M}_K(P)$ , Eq. (5), are still non-zero and  $\Xi^{(Q)}$  is the highest many-body rank of the excited determinant(s)  $|\Phi_K\rangle$  included in  $\mathcal{H}^{(Q)}$ . In practical CC( $P$ ;Q) calculations, including those discussed in Sec. III, the  $\ell_K(P)$  coefficients entering Eq. (8) are calculated as

$$\ell_K(P) = \langle \Phi | (\mathbf{1} + \Lambda^{(P)}) \bar{H}^{(P)} | \Phi_K \rangle / D_K(P), \quad (9)$$

where  $\mathbf{1}$  is the unit operator,

$$\Lambda^{(P)} = \sum_{|\Phi_K\rangle \in \mathcal{H}^{(P)}} \lambda_K (E_K)^\dagger \quad (10)$$

is the hole-particle de-excitation operator defining the bra state  $\langle \tilde{\Psi}^{(P)} | = \langle \Phi | (\mathbf{1} + \Lambda^{(P)}) e^{-T^{(P)}}$  corresponding to the CC( $P$ ) ket state  $|\Psi^{(P)}\rangle = e^{T^{(P)}} |\Phi\rangle$ , and

$$D_K(P) = E^{(P)} - \langle \Phi_K | \bar{H}^{(P)} | \Phi_K \rangle. \quad (11)$$

One determines  $\Lambda^{(P)}$ , or the amplitudes  $\lambda_K$  that define it, by solving the linear system of equations representing the left eigenstate CC problem<sup>10</sup> in the  $P$  space, i.e.,

$$\langle \Phi | (\mathbf{1} + \Lambda^{(P)}) \bar{H}^{(P)} | \Phi_K \rangle = E^{(P)} \lambda_K, \quad |\Phi_K\rangle \in \mathcal{H}^{(P)}, \quad (12)$$

where  $E^{(P)}$  is the previously determined CC( $P$ ) energy. Once the noniterative correction  $\delta(P$ ;Q) is determined, the CC( $P$ ;Q) energy is obtained as

$$E^{(P+Q)} = E^{(P)} + \delta(P;Q). \quad (13)$$

In practice, we often distinguish between the complete version of the CC( $P$ ;Q) theory, designated, following Refs. 58 and 62, as CC( $P$ ;Q)<sub>EN</sub>, which uses the Epstein–Nesbet-like denominator  $D_K(P)$ , Eq. (11), in calculating the  $\ell_K(P)$  amplitudes, and the approximate version of CC( $P$ ;Q), abbreviated as CC( $P$ ;Q)<sub>MP</sub>, which relies on the Møller–Plesset form of  $D_K(P)$  obtained by replacing  $\bar{H}^{(P)}$  in Eq. (11) by the bare Fock operator (see, e.g., Refs. 58, 62, and 63). Both of these variants of the CC( $P$ ;Q) formalism are considered in this study.

We must now come up with the appropriate choices of the  $P$  and  $Q$  spaces entering the CC( $P$ ;Q) considerations that would allow us to match the quality of the high-level CC computations of the CCSDT, CCSDTQ, and similar type at the small fraction of the cost. As is often the case in the CC work, one could start from the conventional choices, where the  $P$  space  $\mathcal{H}^{(P)}$  is spanned by all excited  $|\Phi_{i_1 \dots i_n}^{a_1 \dots a_n}\rangle$  determinants with the excitation rank  $n \leq m_A$ , where  $i_1, i_2, \dots (a_1, a_2, \dots)$  designate the spin-orbitals occupied (unoccupied) in  $|\Phi\rangle$ , and the  $Q$  space  $\mathcal{H}^{(Q)}$  by those with  $m_A < n \leq m_B$ , where  $m_B \leq N$ . In that case, one ends up with the well-established CR-CC( $m_A, m_B$ ) hierarchy,<sup>52,57,63,75–80,83</sup> including the aforementioned CR-CC(2,3) approximation, where  $m_A = 2$  and  $m_B = 3$ , and the related CCSD(2)<sub>T</sub><sup>88</sup> (see also Refs. 89–91), CCSD(T)<sub>Λ</sub>,<sup>92–94</sup> Λ-CCSD(T),<sup>95,96</sup> and similar<sup>46,47,86,87,97</sup> schemes that allow one to correct the CCSD energies for triples. The CR-CC(2,3) method is useful, improving, for example, poor performance of

CCSD(T) in covalent bond breaking situations<sup>57,75–78,83,98,99</sup> and for certain classes of noncovalent interactions<sup>82,100</sup> without a substantial increase in the computational effort, but neither CR-CC(2,3) nor its CCSD(2)<sub>T</sub>, CCSD(T)<sub>Λ</sub>, and Λ-CCSD(T) counterparts [which are all approximations to CR-CC(2,3)] are free from drawbacks. One of the main problems with CR-CC(2,3), CCSD(2)<sub>T</sub>, Λ-CCSD(T), and other noniterative corrections to CCSD is the fact that, in analogy to CCSD(T), they decouple the higher-order  $T_n$  components with  $n > m_A$ , such as  $T_3$  or  $T_3$  and  $T_4$ , from their lower-order  $n \leq m_A$  (e.g.,  $T_1$  and  $T_2$ ) counterparts. This can result in substantial errors, for example, when the activation energies and chemical reaction profiles involving rearrangements of  $\pi$  bonds and singlet–triplet gaps in certain classes of biradical species are examined.<sup>61–63</sup> The automerization of cyclobutadiene, which is one of the benchmark examples in Sec. III, provides an illustration of the challenges the noniterative corrections to CCSD, including CCSD(2)<sub>T</sub> and CR-CC(2,3), face when the coupling of the lower-order  $T_1$  and  $T_2$  and higher-order  $T_3$  clusters becomes significant (see Ref. 61 for further analysis and additional remarks). One can address problems of this type by using active orbitals to incorporate the dominant higher-than-doubly excited determinants, in addition to all singles and doubles, in the  $P$  space, as in the successful CC(t;3), CC(t,q;3), and CC(t,q;3,4) hierarchy,<sup>57,61–63,82,100</sup> which uses the CC( $P$ ;Q) framework to correct the results of the active-space CCSDt<sup>48,101–108</sup> or CCSDtq<sup>48,102,103,107,109</sup> calculations for the remaining  $T_3$  or  $T_3$  and  $T_4$  correlations that were not captured via active orbitals, but the resulting methods are no longer computational black boxes. The semi-stochastic CC( $P$ ;Q) methodology, introduced in Ref. 58, extended to excited states in Ref. 60, and further developed in this work, which takes advantage of the FCIQMC or truncated CIQMC/CCMC propagations that can identify the leading higher-than-doubly excited determinants for the inclusion in the  $P$  space, while using the noniterative  $\delta(P$ ;Q) corrections to capture the remaining correlations of interest, offers an automated way of performing CC( $P$ ;Q) computations without any reference to the user- and system-dependent active orbitals. The semi-stochastic CC( $P$ ;Q) methods developed and tested in this study are discussed next.

## B. Semi-stochastic CC( $P$ ;Q) approaches using FCIQMC and its truncated CISDT-MC and CISDTQ-MC counterparts

In our original examination of the semi-stochastic CC( $P$ ;Q) framework<sup>58</sup> and its recent extension to excited states,<sup>60</sup> where we focused on converging the full CCSDT and EOMCCSDT energetics, we demonstrated that the FCIQMC and CCSDT-MC approaches are capable of generating meaningful  $P$  spaces for the subsequent CC( $P$ )/EOMCC( $P$ ) iterations, which precede the determination of the  $\delta(P$ ;Q) moment corrections, already in the early stages of the respective QMC propagations. The main objective of this work is to explore if the same remains true when FCIQMC is replaced by its less expensive truncated CISDT-MC and CISDTQ-MC counterparts, in which spawning beyond the triply excited (CISDT-MC) or quadruply excited (CISDTQ-MC) determinants is disallowed, and if one can use the CIQMC-driven CC( $P$ ;Q) calculations to converge the higher-level CCSDTQ energetics with similar efficiency.

The key steps of the semi-stochastic CC( $P$ ;Q) algorithm exploited in this study, which allows us to converge the CCSDT and CCSDTQ energetics using the  $P$  spaces extracted from the FCIQMC and truncated CISDT-MC and CISDTQ-MC propagations, are as follows:

1. Initiate a CIQMC run appropriate for the CC method of interest by placing a certain number of walkers on the reference state  $|\Phi\rangle$ , which in all of the calculations reported in this article is the restricted Hartree–Fock (RHF) determinant. Among the CIQMC schemes that can provide meaningful  $P$  spaces for the CC( $P$ ;Q) calculations targeting the CCSDT energetics are the FCIQMC approach used in our earlier work<sup>58–60</sup> and the CISDT-MC and CISDTQ-MC methods examined in the present study. If the objective is to converge the CCSDTQ energetics, one can use FCIQMC or CISDTQ-MC, which are the two choices pursued in the present work, but not CISDT-MC, which ignores quadruply excited determinants. As in our earlier semi-stochastic CC( $P$ )/EOMCC( $P$ ) and CC( $P$ ;Q) work,<sup>58–60</sup> all of the calculations reported in this article adopted the initiator CIQMC ( $i$ -CIQMC) algorithm, originally proposed in Ref. 65, based on integer walker numbers, but the procedure discussed here is flexible and could be merged with other CIQMC techniques developed in recent years, such as those described in Refs. 67 and 114.
2. After a certain number of CIQMC time steps, called MC iterations, i.e., after some QMC propagation time  $\tau$ , extract a list of higher-than-doubly excited determinants relevant to the CC theory of interest to construct the  $P$  space for executing the CC( $P$ ) calculations. If one is interested in targeting the CCSDT-level energetics, the  $P$  space used in the CC( $P$ ) iterations consists of all singly and doubly excited determinants and a subset of triply excited determinants identified by the underlying FCIQMC, CISDT-MC, or CISDTQ-MC propagation, where each triply excited determinant in the subset is populated by at least  $n_p$  positive or negative walkers. In analogy to our previous CC( $P$ )/EOMCC( $P$ ) and CC( $P$ ;Q) studies,<sup>58–60</sup> all of the CC( $P$ ) and CC( $P$ ;Q) computations carried out in this work use  $n_p = 1$ . If the goal is to converge the CCSDTQ energetics, the  $P$  space for the CC( $P$ ) computations is defined as all singly and doubly excited determinants and a subset of triply and quadruply excited determinants identified by the underlying FCIQMC or CISDTQ-MC propagation, where, again, each triply and quadruply excited determinant in the subset is populated by a minimum of  $n_p$  positive or negative walkers.
3. Solve the CC( $P$ ) and left-eigenstate CC( $P$ ) equations, Eqs. (4) and (12), respectively, where  $E^{(P)}$  is given by Eq. (7), for the cluster operator  $T^{(P)}$  and the de-excitation operator  $\Lambda^{(P)}$  in the  $P$  space determined in step 2. If the objective is to converge the CCSDT-level energetics, we define  $T^{(P)} = T_1 + T_2 + T_3^{(\text{MC})}$  and  $\Lambda^{(P)} = \Lambda_1 + \Lambda_2 + \Lambda_3^{(\text{MC})}$ , where  $T_3^{(\text{MC})}$  and  $\Lambda_3^{(\text{MC})}$  are the three-body components of  $T^{(P)}$  and  $\Lambda^{(P)}$ , respectively, defined using the list of triples identified by the FCIQMC, CISDT-MC, or CISDTQ-MC propagation at time  $\tau$ , as described in step 2.

- If one is targeting the CCSDTQ-level energetics,  $T^{(P)} = T_1 + T_2 + T_3^{(\text{MC})} + T_4^{(\text{MC})}$  and  $\Lambda^{(P)} = \Lambda_1 + \Lambda_2 + \Lambda_3^{(\text{MC})} + \Lambda_4^{(\text{MC})}$ , where  $T_3^{(\text{MC})}$  and  $\Lambda_3^{(\text{MC})}$  are the three-body components and  $T_4^{(\text{MC})}$  and  $\Lambda_4^{(\text{MC})}$  are the four-body components of  $T^{(P)}$  and  $\Lambda^{(P)}$ , respectively, defined using the lists of triples and quadruples identified by the FCIQMC or CISDTQ-MC propagation at time  $\tau$ .
- Use the CC( $P$ ;Q) correction  $\delta(P;Q)$ , Eq. (8), to correct the energy  $E^{(P)}$  obtained in step 3 for the remaining correlation effects of interest, meaning those correlations that were not captured by the CC( $P$ ) calculations performed at the time  $\tau$  the list of higher-than-doubly excited determinants entering the relevant  $P$  space was created. If the objective is to converge the CCSDT-level energetics, the  $Q$  space entering the definition of  $\delta(P;Q)$  consists of those triply excited determinants that in the FCIQMC, CISDT-MC, or CISDTQ-MC propagation at time  $\tau$  are populated by less than  $n_p$  positive or negative walkers (in this study, where  $n_p = 1$ , the triply excited determinants that were not captured by the FCIQMC, CISDT-MC, or CISDTQ-MC propagation at time  $\tau$ ). If the goal is to recover the CCSDTQ-level energetics, the  $Q$  space used to calculate  $\delta(P;Q)$  consists of the triply and quadruply excited determinants that in the FCIQMC or CISDTQ-MC propagation at time  $\tau$  are populated by less than  $n_p$  positive or negative walkers.
  - Check the convergence of the CC( $P$ ;Q) energy  $E^{(P+Q)}$ , Eq. (13), obtained in step 4 by repeating steps 2–4 at some later CIQMC propagation time  $\tau' > \tau$ . If the resulting energy  $E^{(P+Q)}$  no longer changes within a given convergence threshold, the CC( $P$ ;Q) calculation can be stopped. As pointed out in Refs. 58–60, one can also stop it once the fraction (fractions) of higher-than-doubly excited determinants captured by the CIQMC propagation relevant to the target CC theory level, included in the  $P$  space, is (are) sufficiently large to obtain the desired accuracy. This is further discussed in Sec. III, where the numerical results obtained in this study are presented.

The above semi-stochastic CC( $P$ ;Q) algorithm, allowing us to recover the CCSDT and CCSDTQ energetics using the  $P$  spaces identified with the help of FCIQMC or truncated CISDT-MC and CISDTQ-MC propagations, has been implemented by modifying our previously developed standalone deterministic CC( $P$ ;Q) codes,<sup>57,61,63</sup> which rely on the RHF, restricted open-shell Hartree-Fock, and integral routines in the GAMESS package,<sup>110,111</sup> such that they could handle the stochastically determined lists of triples and quadruples, and by interfacing the resulting program with the  $i$ -CIQMC routines available in the HANDE software.<sup>112,113</sup> As in our earlier semi-stochastic CC( $P$ )/EOMCC( $P$ ) and CC( $P$ ;Q) work,<sup>58–60</sup> we rely on the original form of the initiator CIQMC ( $i$ -CIQMC) algorithm proposed in Ref. 65, where only those determinants that acquire walker population exceeding a preset value  $n_a$  are allowed to attempt spawning new walkers onto empty determinants, but, as already alluded to above, one could consider interfacing our CC( $P$ ;Q) framework with the improved ways of converging CIQMC, such as the adaptive-shift method developed in Refs. 67 and 114. While the choice of a specific CIQMC algorithm may not be as critical in the context of CC( $P$ ;Q) considerations as in the case

of other applications of QMC techniques, since the only role of CIQMC propagations in the semi-stochastic CC( $P$ ;Q) calculations is to identify the leading higher-than-doubly excited determinants for the inclusion in the  $P$  space and, as shown in Sec. III and our previous studies,<sup>58,60</sup> moment corrections  $\delta(P;Q)$  are very efficient in accounting for the many-electron correlation effects due to the remaining determinants not captured by CIQMC, we are planning to integrate our CC( $P$ ;Q) codes with the CIQMC methods described in Refs. 67 and 114 in the future work. It will be interesting to examine if the excellent performance of the semi-stochastic CC( $P$ ;Q) methods observed in the calculations reported in this article can be improved further by replacing the  $i$ -CIQMC algorithm by better ways of converging CIQMC.

In the case of the semi-stochastic CC( $P$ ;Q) codes aimed at converging the CCSDT energetics, which we have extended in the present study by allowing them to work with the CISDT-MC and CISDTQ-MC approaches, in addition to the previously examined FCIQMC<sup>58–60</sup> and CCSDT-MC<sup>58</sup> options, we follow the algorithm summarized in steps 1–5 without any alterations. In particular, all of the quantities entering Eq. (8) for the noniterative correction  $\delta(P;Q)$  are treated in the present study fully. This is an improvement compared to our original semi-stochastic CC( $P$ ;Q) computations utilizing FCIQMC and CCSDT-MC, reported in Ref. 58, where we adopted an approximation in which the three-body component  $\Lambda_3^{(\text{MC})}$  of the de-excitation operator  $\Lambda^{(P)}$  used to determine amplitudes  $\ell_K(P)$  entering Eq. (8) was neglected. In analogy to this work, the similarity-transformed Hamiltonian  $\overline{H}^{(P)}$ , defining moments  $\mathfrak{M}_K(P)$  and entering the linear system defined by Eq. (12), which is used to determine  $\Lambda^{(P)}$ , was treated in Ref. 58 fully, i.e.,  $\overline{H}^{(P)}$  employed in the CC( $P$ ;Q) calculations aimed at recovering the CCSDT energetics was defined as  $(He^{T_1+T_2+T_3^{(\text{MC})}})_C$ , so that the one- and two-body components of  $\Lambda^{(P)}$  employed in Ref. 58 were properly relaxed in the presence of the three-body component  $T_3^{(\text{MC})}$  of the cluster operator  $T^{(P)}$  obtained in the preceding CC( $P$ ) calculations, but  $\Lambda_3^{(\text{MC})}$  was neglected. Although all of our numerical tests to date indicate that this approximation has a small effect on the results of the semi-stochastic CC( $P$ ;Q) calculations utilizing full and truncated CIQMC and no effect on our main conclusions, we no longer use it in this work. In other words, all of the calculations reported in the present study rely on the complete representations of  $\overline{H}^{(P)}$  and  $\Lambda^{(P)}$  when constructing moments  $\mathfrak{M}_K(P)$  and amplitudes  $\ell_K(P)$  entering Eq. (8). This means that  $\overline{H}^{(P)}$  and  $\Lambda^{(P)}$  used to determine the CC( $P$ ;Q) correction  $\delta(P;Q)$  in the calculations aimed at the CCSDT energetics are defined as  $(He^{T_1+T_2+T_3^{(\text{MC})}})_C$  and  $\Lambda_1 + \Lambda_2 + \Lambda_3^{(\text{MC})}$ , respectively.

We have, however, introduced an approximation in the semi-stochastic CC( $P$ ;Q) routines that are used to converge the CCSDTQ-level energetics. Given the pilot nature of these routines, the noniterative correction  $\delta(P;Q)$  that they produce corrects the  $E^{(P)}$  energy, which is obtained in this case by solving the CC( $P$ ) equations in the space of all singles and doubles and subsets of triples and quadruples captured by FCIQMC or CISDTQ-MC, for the remaining triples not included in the  $P$  space, but the quadruples contributions to  $\delta(P;Q)$  are ignored. This approximation is acceptable since in the  $\tau = \infty$  limit, where the  $P$  space contains all triples and quadruples, i.e., the corresponding  $Q$  space is empty, the uncorrected CC( $P$ ) and

partially or fully corrected  $CC(P;Q)$  calculations recover the CCSDTQ energetics. All of our tests to date, including those discussed in Sec. III, indicate that the convergence of the  $CC(P;Q)$  computations, in which the quadruples component of  $\delta(P;Q)$  is ignored, toward CCSDTQ is rapid, even when the  $T_4$  effects become significant, so the above approximation does not seem to have a major effect on the convergence rate, but we will implement the full correction  $\delta(P;Q)$  due to the missing triples as well as quadruples in the future to examine if one can accelerate convergence toward CCSDTQ even further.

As explained in Refs. 58 and 60 (see also Ref. 59), the semi-stochastic  $CC(P;Q)$  approaches of the type summarized above offer a number of advantages. Among them are substantial savings in the computational effort compared to the parent high-level CC theories they target and a systematic behavior of the resulting  $E^{(P+Q)}$  energies as  $\tau$  approaches  $\infty$ . The latter feature is a direct consequence of the fact that if we follow the definitions of the  $P$  and  $Q$  spaces introduced in steps 2 and 4 above, the initial,  $\tau = 0$ ,  $CC(P;Q)$  energies are identical to those obtained with CR-CC(2,3) or CR-CC(2,4), which are approximations to CCSDT and CCSDTQ, respectively, that account for some  $T_3$  [CR-CC(2,3)] or  $T_3$  and  $T_4$  [CR-CC(2,4)] correlations. In the  $\tau = \infty$  limit, the  $CC(P;Q)$  energies  $E^{(P+Q)}$  become equivalent to their respective high-level CC parents, which account for the  $T_n$  components with  $n > 2$ , such as  $T_3$  or  $T_3$  and  $T_4$ , fully, so that the QMC propagation time  $\tau$  becomes a parameter connecting CR-CC(2,3) with CCSDT and CR-CC(2,4) with CCSDTQ. In the case of our current implementation of the semi-stochastic  $CC(P;Q)$  approach aimed at converging the CCSDTQ energetics, where the quadruples contributions to correction  $\delta(P;Q)$  are ignored, the initial,  $\tau = 0$ ,  $CC(P;Q)$  energy is equivalent to that obtained with the CR-CC(2,3) approach, i.e., the QMC propagation time  $\tau$  connects CR-CC(2,3) with CCSDTQ. When  $\tau$  approaches  $\infty$ , the uncorrected  $CC(P)$  energies  $E^{(P)}$  converge to their CCSDT and CCSDTQ parents as well, but the convergence toward CCSDT and CCSDTQ is in this case slower, since the  $CC(P)$  energies at  $\tau = 0$  are equivalent to those of CCSD, which has no information about the  $T_n$  components with  $n > 2$ , and, as shown in our earlier work,<sup>58,60</sup> and as clearly demonstrated in the present study, the  $CC(P;Q)$  corrections  $\delta(P;Q)$  greatly accelerate the convergence toward the target CC energetics. The above relationships between the semi-stochastic  $CC(P)$  and  $CC(P;Q)$  approaches and the deterministic CCSD, CR-CC(2,3)/CR-CC(2,4), and CCSDT/CCSDTQ theories are also helpful when debugging the  $CC(P)$  and  $CC(P;Q)$  codes.

As far as the savings in the computational effort offered by the semi-stochastic  $CC(P;Q)$  methods, when compared to their high-level CC parents, such as CCSDT or CCSDTQ, are concerned, they were already discussed in Refs. 58 and 60, so here we focus on the information relevant to the calculations discussed in Sec. III. There are three main factors that contribute to these savings. First, the computational times associated with the early stages of the CIQMC walker propagations, which are sufficient to recover the parent CCSDT or CCSDTQ energetics to within small fractions of a millihartree when the semi-stochastic  $CC(P;Q)$  framework is employed, are very short compared to the converged CIQMC runs. They are already short when one uses FCIQMC, and they are even shorter when one replaces FCIQMC by the CISDT-MC and CISDTQ-MC truncations. As further elaborated in Sec. III, the convergence of the semi-stochastic  $CC(P;Q)$  calculations toward the parent CCSDT and

CCSDTQ energies is so fast that the underlying CIQMC computations use much smaller walker populations than those required to converge the CIQMC propagations. They are small when one uses FCIQMC, and they become even smaller when one relies on the truncated CISDT-MC and CISDTQ-MC approaches in the  $CC(P;Q)$  runs.

Second, the  $CC(P)$  calculations using small fractions of higher-than-doubly excited determinants, which is how the  $P$  spaces used in these calculations look like when the early stages of the CIQMC walker propagations are considered, are much faster than the parent CC computations. For example, when the most expensive  $\langle \Phi_{ijk}^{abc} | [H, T_3] | \Phi \rangle$  or  $\langle \Phi_{ijk}^{abc} | [\bar{H}^{(2)}, T_3] | \Phi \rangle$  contributions to the CCSDT equations, where  $\bar{H}^{(2)} = e^{-T_1 - T_2} H e^{T_1 + T_2}$ , are isolated and implemented using programming methods similar to those exploited in selected CI algorithms (rather than the usual diagrammatic techniques that assume continuous excitation manifolds labeled by all occupied and all unoccupied orbitals), one can accelerate their determination by a factor of up to  $(D/d)^2$ , where  $D$  is the number of all triples and  $d$  is the number of triples included in the  $P$  space, captured with the help of CIQMC propagations. Other contributions to the CCSDT equations that involve  $T_3$  or the projections on the triply excited determinants, such as  $\langle \Phi_{ij}^{ab} | [H, T_3] | \Phi \rangle$  and  $\langle \Phi_{ijk}^{abc} | [H, T_2] | \Phi \rangle$ , may offer additional speedups, on the order of  $(D/d)$ . Our current  $CC(P)$  codes are still in the pilot stages, but the speedups on the order of  $(D/d)$  in the determination of the most expensive  $\langle \Phi_{ijk}^{abc} | [H, T_3] | \Phi \rangle$  (or  $\langle \Phi_{ijk}^{abc} | [\bar{H}^{(2)}, T_3] | \Phi \rangle$ ) terms are attainable. Similar remarks apply to the  $CC(P)/CC(P;Q)$  calculations aimed at converging the CCSDTQ energetics, where one can considerably speed up the determination of the most expensive  $\langle \Phi_{ijkl}^{abcd} | [H, T_4] | \Phi \rangle$  or  $\langle \Phi_{ijkl}^{abcd} | [\bar{H}^{(2)}, T_4] | \Phi \rangle$  contributions and other terms containing the  $T_3$  and  $T_4$  clusters and the projections on the triply and quadruply excited determinants. It should also be noted that the  $CC(P)$  calculations do not require storing the entire  $T_3$  and  $T_4$  vectors. The  $T_3^{(MC)}$  and  $T_4^{(MC)}$  operators use much smaller numbers of amplitudes than their full  $T_3$  and  $T_4$  counterparts.

Third, the computation of the noniterative correction  $\delta(P;Q)$  is much less expensive than a single iteration of the target CC calculation. In the case of the  $CC(P;Q)$  calculations aimed at converging the CCSDT energetics, the computational time required to determine the corresponding correction  $\delta(P;Q)$  scales no worse than  $\sim 2n_o^3 n_u^4$ , which is much less than the  $n_o^3 n_u^5$  scaling of each iteration of CCSDT. In the case of the  $CC(P;Q)$  approach aimed at CCSDTQ, the computational time required to determine correction  $\delta(P;Q)$  scales as  $\sim 2n_o^3 n_u^4$  in the case of the contributions due to the remaining triples and is on the order of  $n_o^4 n_u^5$  in the case of the quadruples part of  $\delta(P;Q)$ , when the more complete  $CC(P;Q)_{EN}$  approach is used, or  $n_o^2 n_u^5$ , when the  $CC(P;Q)_{MP}$  form of  $\delta(P;Q)$  is employed. This is all much less than the  $n_o^4 n_u^6$  scaling of every CCSDTQ iteration. As mentioned above, in our current implementation of the semi-stochastic  $CC(P;Q)$  approach aimed at converging the CCSDTQ energetics, the quadruples contribution to correction  $\delta(P;Q)$  is neglected, so the computational time required to obtain  $\delta(P;Q)$  scales as  $\sim 2n_o^3 n_u^4$ , at worst, which points to the usefulness of such an approximation, especially that the convergence of the resulting  $CC(P;Q)$  energies toward CCSDTQ is, as shown in Sec. III, very fast.

### III. NUMERICAL EXAMPLES

In order to demonstrate the benefits offered by the semi-stochastic  $CC(P;Q)$  framework, especially the new  $CC(P;Q)$  approaches implemented in this work that replace FCIQMC by the less expensive CISDT-MC and CISDTQ-MC propagations, we applied the FCIQMC-, CISDT-MC-, and CISDTQ-MC-driven  $CC(P;Q)$  methods aimed at converging the CCSDT and CCSDTQ energetics to a few molecular problems, for which the parent full CCSDT and CCSDTQ results had previously been determined or were not too difficult to be recalculated. Thus, we carried out an extensive series of the CISDT-MC- and CISDTQ-MC-driven  $CC(P;Q)$  calculations, along with the analogous computations using FCIQMC, which was utilized in our earlier study,<sup>58</sup> to examine the ability of the semi-stochastic  $CC(P;Q)$  approaches using various types of CIQMC to recover the CCSDT energetics for the F–F bond dissociation in the fluorine molecule (Sec. III A) and the automerization of cyclobutadiene (Sec. III B). In order to illustrate the performance of the FCIQMC- and CISDTQ-MC-driven  $CC(P;Q)$  methods in calculations aimed at converging the CCSDTQ energetics, we considered the symmetric stretching of the O–H bonds in the water molecule (Sec. III C). We chose bond breaking in  $F_2$ , which is accurately described by full CCSDT,<sup>57,61,75,76,115</sup> since we examined the same system in our original FCIQMC- and CCSDT-MC-driven  $CC(P;Q)$  work<sup>58</sup> and in the preceding deterministic  $CC(P;Q)$ -based  $CC(t;3)$  calculations reported in Ref. 57. Our choice of the automerization of cyclobutadiene, which is accurately described by CCSDT as well,<sup>61,116</sup> was motivated by similar reasons. We studied this problem, where all noniterative triples corrections to CCSD, including CCSD(T),  $\Lambda$ -CCSD(T), CCSD(2)<sub>T</sub>, and CR-CC(2,3) fail,<sup>61,116,117</sup> using the deterministic  $CC(t;3)$  approach exploiting the  $CC(P;Q)$  ideas in Ref. 61, and we studied it again using the semi-stochastic  $CC(P;Q)$  framework utilizing FCIQMC and CCSDT-MC in Ref. 58. We would like to explore now what the effect of replacing FCIQMC propagations by their less expensive CISDT-MC and CISDTQ-MC counterparts on the convergence of the  $CC(P;Q)$  energies toward CCSDT is. We would also like to learn if the incorporation of the previously neglected<sup>58</sup> three-body component of the de-excitation operator  $\Lambda^{(P)}$ , which is used to construct amplitudes  $\ell_K(P)$  entering Eq. (8), helps the accuracy of the resulting semi-stochastic  $CC(P;Q)$  energies. We studied the  $C_{2v}$ -symmetric double dissociation of  $H_2O$ , since by simultaneously stretching both O–H bonds by factors exceeding 2, one ends up with a catastrophic failure of CCSDT.<sup>63,75,118</sup> One needs an accurate description of the  $T_3$  and  $T_4$  clusters to obtain a more reliable description of the water potential energy surface in that region.<sup>63</sup>

Following our earlier semi-stochastic and deterministic  $CC(P;Q)$  work,<sup>57,58,61,63</sup> which also provides the parent CCSDT<sup>57,58,61,63</sup> and CCSDTQ<sup>63</sup> energetics, we used the cc-pVDZ,<sup>119</sup> cc-pVTZ,<sup>119</sup> and aug-cc-pVTZ<sup>120</sup> basis sets for  $F_2$  and the cc-pVDZ bases for cyclobutadiene and water. For consistency with Refs. 57, 58, and 61, in all of the post-RHF computations for the F–F bond breaking in  $F_2$  and the automerization of cyclobutadiene, the core electrons corresponding to the 1s shells of the fluorine and carbon atoms were kept frozen. As in Refs. 63 and 118, which provide the reference CCSDTQ data and, in the case of Ref. 118, the geometries of the equilibrium and stretched water molecule used in our semi-stochastic  $CC(P;Q)$  calculations aimed at converging the CCSDTQ energetics, we correlated

all electrons. Each of the relevant  $i$ -FCIQMC (all systems),  $i$ -CISDT-MC ( $F_2$  and cyclobutadiene), and  $i$ -CISDTQ-MC (all systems) runs was initiated by placing 100 walkers on the RHF reference determinant, and we set the initiator parameter  $n_a$  at 3. All of the  $i$ -FCIQMC,  $i$ -CISDT-MC, and  $i$ -CISDTQ-MC propagations used the time step  $\delta\tau$  of 0.0001 a.u.

#### A. Bond breaking in $F_2$

We begin our discussion of the semi-stochastic  $CC(P;Q)$  calculations carried out in this study with the F–F bond dissociation in the fluorine molecule, as described by the cc-pVDZ basis set using the Cartesian components of  $d$  orbitals (see Table I and Figs. 1–3). In analogy to Ref. 58, where our initial FCIQMC- and CCSDT-MC-based  $CC(P;Q)$  results for  $F_2$  were presented, we considered the equilibrium geometry  $R_e = 2.66816$  bohr, where the many-electron correlation effects have a predominantly dynamical character, and three stretches of the F–F bond length  $R$ , including  $R = 1.5R_e$ ,  $2R_e$ , and  $5R_e$ , which are characterized by the increasingly large nondynamical correlations. The increasingly important role of nondynamical correlation effects as the F–F bond is stretched is reflected in the magnitude of  $T_3$  contributions, defined by forming the difference of the CCSDT and CCSD energies, which grows, in absolute value, from 9.485 millihartree at  $R = R_e$  to 32.424, 45.638, and 49.816 millihartree at  $R = 1.5R_e$ ,  $2R_e$ , and  $5R_e$ , respectively, when the cc-pVDZ basis set is employed. The  $T_3$  effects in the  $R = 2R_e - 5R_e$  region are so large that they exceed the depth of the CCSDT potential well, estimated at about 44 millihartree when the difference between the CCSDT energies at  $R = 5R_e$ , where  $F_2$  is essentially dissociated, and  $R = R_e$  is considered. They grow with  $R$  so fast that the popular perturbative CCSD(T) correction to CCSD fails at larger F–F separations, producing the  $-5.711$ ,  $-23.596$ , and  $-39.348$  millihartree errors relative to CCSDT at  $R = 1.5R_e$ ,  $2R_e$ , and  $5R_e$ , respectively, misrepresenting the physics of  $T_3$  correlations in the stretched  $F_2$  molecule.

The triples corrections to CCSD that rely on the biorthogonal moment expansions of the  $CC(P;Q)$  type, including CR-CC(2,3), work much better than CCSD(T). This is especially true when the most complete variant of the CR-CC(2,3) approach using the Epstein–Nesbet form of the  $D_K(P)$  denominator in determining the  $\ell_K(P)$  amplitudes that enter the corresponding triples correction to CCSD, abbreviated sometimes as CR-CC(2,3)<sub>D</sub> or CR-CC(2,3)<sub>D</sub><sup>62,63,78,80,83</sup> and represented in Table I by the  $\tau = 0$   $CC(P;Q)_{EN}$  results, is considered. Indeed, the CR-CC(2,3)<sub>D</sub> calculations reduce large errors in the CCSD(T) energies at  $R = 1.5R_e$ ,  $2R_e$ , and  $5R_e$  to 1.735, 1.862, and 1.613 millihartree, respectively, improving the CCSD(2)<sub>T</sub> or the equivalent<sup>62,63,78,80,83</sup> CR-CC(2,3)<sub>A</sub> or CR-CC(2,3)<sub>A</sub> calculations, which adopt the Møller–Plesset  $D_K(P)$  denominators, at the same time [see the  $\tau = 0$   $CC(P;Q)_{MP}$  values in Table I]. The CR-CC(2,3)<sub>D</sub> approach eliminates the failure of CCSD(T) at stretched nuclear geometries while being more effective in capturing the physics of  $T_3$  correlations than CCSD(2)<sub>T</sub>, but the only way to obtain further improvements toward CCSDT is by incorporating at least some triples in the iterative part of the calculations, relaxing the  $T_1$  and  $T_2$  amplitudes, which in CCSD(T), CCSD(2)<sub>T</sub>, and CR-CC(2,3) are fixed at their CCSD values, in the presence of the leading  $T_3$  contributions, and correcting the resulting energies for the remaining  $T_3$  effects accordingly. One can



**TABLE I.** 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  $F_2/cc$ -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			$CC(P)^a$			$CC(P;Q)_{MP}^a$			$CC(P;Q)_{EN}^a$		
		FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>
1.0	0		0			9.485 <sup>e</sup>			1.398 <sup>f</sup>			−0.240 <sup>g</sup>	
	10 000	3	3	4	5.692	5.692	5.229	0.760	0.760	0.688	−0.151	−0.151	−0.152
	20 000	9	8	8	3.548	3.804	3.962	0.444	0.473	0.472	−0.107	−0.093	−0.140
	30 000	15	16	14	2.290	2.498	2.769	0.284	0.301	0.334	−0.059	−0.046	−0.067
	40 000	25	26	22	1.791	1.523	1.765	0.212	0.184	0.210	−0.037	−0.030	−0.034
	50 000	37	38	34	0.933	0.940	1.151	0.113	0.115	0.137	−0.014	−0.013	−0.021
	60 000	51	52	46	0.536	0.498	0.698	0.064	0.058	0.083	−0.008	−0.008	−0.010
	70 000	63	64	58	0.383	0.308	0.410	0.044	0.036	0.047	−0.006	−0.004	−0.007
	80 000	73	74	68	0.177	0.164	0.224	0.020	0.018	0.025	−0.003	−0.002	−0.003
	100 000	89	89	85	0.044	0.050	0.073	0.005	0.006	0.008	0.000	−0.001	−0.001
120 000	97	97	94	0.013	0.010	0.024	0.001	0.001	0.003	0.000	0.000	0.000	
	∞		100		−199.102 796 <sup>h</sup>			...			...		
1.5	0		0			32.424 <sup>e</sup>			5.984 <sup>f</sup>			1.735 <sup>g</sup>	
	10 000	3	3	3	14.312	14.220	15.874	2.198	1.980	2.115	0.351	0.321	0.193
	20 000	9	8	7	5.589	3.572	5.564	0.629	0.428	0.657	−0.003	−0.000	0.052
	30 000	16	18	14	2.728	2.391	2.206	0.323	0.285	0.262	−0.002	0.020	0.021
	40 000	27	30	24	1.065	0.706	1.387	0.142	0.084	0.171	0.020	0.009	0.015
	50 000	42	45	35	0.482	0.459	0.687	0.062	0.055	0.087	0.009	0.006	0.008
	60 000	57	60	49	0.273	0.219	0.336	0.029	0.027	0.041	0.001	0.000	0.005
	70 000	70	72	61	0.128	0.106	0.231	0.013	0.011	0.028	0.000	0.000	0.001
	80 000	81	82	72	0.064	0.048	0.102	0.006	0.004	0.010	0.000	−0.001	−0.001
	100 000	93	94	88	0.012	0.009	0.026	0.001	0.001	0.003	0.000	0.000	0.000
120 000	99	100	96	0.001	0.002	0.005	0.000	0.000	0.000	0.000	0.000	0.000	
	∞		100		−199.065 882 <sup>h</sup>			...			...		
2.0	0		0			45.638 <sup>e</sup>			6.357 <sup>f</sup>			1.862 <sup>g</sup>	
	10 000	4	4	3	12.199	17.779	12.687	0.998	1.886	1.181	−0.063	0.280	−0.008
	20 000	10	9	9	4.127	2.529	3.672	0.328	0.245	0.310	−0.014	0.009	−0.025
	30 000	21	19	17	0.802	1.172	1.393	0.081	0.115	0.128	0.008	0.011	0.004
	40 000	35	32	28	0.456	0.499	0.627	0.040	0.047	0.058	−0.001	0.000	0.000
	50 000	51	48	41	0.216	0.215	0.305	0.018	0.019	0.027	−0.001	0.000	−0.001
	60 000	66	64	56	0.083	0.112	0.160	0.007	0.010	0.014	−0.001	−0.001	−0.001
	70 000	79	75	68	0.037	0.048	0.074	0.003	0.004	0.006	0.000	−0.001	−0.001
	80 000	87	85	78	0.013	0.019	0.034	0.001	0.002	0.003	0.000	0.000	0.000
	100 000	97	95	91	0.001	0.002	0.007	0.000	0.000	0.001	0.000	0.000	0.000
120 000	100	100	98	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	∞		100		−199.058 201 <sup>h</sup>			...			...		

TABLE I. (Continued.)

$R/R_e$	MC iterations	% of triples			CC( $P$ ) <sup>a</sup>			CC( $P;Q$ ) <sub>MP</sub> <sup>a</sup>			CC( $P;Q$ ) <sub>EN</sub> <sup>a</sup>		
		FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>
5.0	0		0			49.816 <sup>e</sup>		3.895 <sup>f</sup>			1.613 <sup>g</sup>		
	10 000	3	3	3	10.887	13.326	9.776	0.455	0.672	0.642	−0.005	0.059	0.202
	20 000	8	8	8	1.968	2.535	1.315	0.152	0.165	0.102	0.040	0.026	0.012
	30 000	17	15	15	0.529	0.752	1.042	0.041	0.056	0.081	0.001	0.006	0.015
	40 000	27	26	26	0.295	0.351	0.346	0.022	0.024	0.025	0.001	−0.001	−0.001
	50 000	38	37	36	0.116	0.147	0.166	0.008	0.011	0.011	−0.001	0.000	−0.001
	60 000	47	46	44	0.047	0.059	0.070	0.003	0.004	0.005	−0.001	0.000	−0.001
	70 000	54	52	50	0.016	0.020	0.030	0.001	0.001	0.002	0.000	0.000	0.000
	80 000	60	59	55	0.006	0.006	0.014	0.000	0.000	0.001	0.000	0.000	0.000
	100 000	74	73	66	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	120 000	89	87	78	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	∞		100			−199.058 586 <sup>h</sup>			...			...	

<sup>a</sup>Unless otherwise stated, all energies are reported as errors relative to CCSDT in millihartree.

<sup>b</sup>FCI stands for *i*-FCIQMC.

<sup>c</sup>CIQ stands for *i*-CISDTQ-MC.

<sup>d</sup>CIT stands for *i*-CISDT-MC.

<sup>e</sup>Equivalent to CCSD.

<sup>f</sup>Equivalent to the CCSD energy corrected for the effects of  $T_3$  clusters using the CCSD(2)<sub>T</sub> approach of Ref. 88, which is equivalent to the approximate form of the completely renormalized CR-CC(2,3) approach of Refs. 75 and 76, abbreviated sometimes as CR-CC(2,3)<sub>A</sub> or CR-CC(2,3)<sub>A</sub>.<sup>62,63,78,80,83</sup>

<sup>g</sup>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. 75 and 76, abbreviated sometimes as CR-CC(2,3)<sub>D</sub> or CR-CC(2,3)<sub>D</sub>.<sup>62,63,78,80,83</sup>

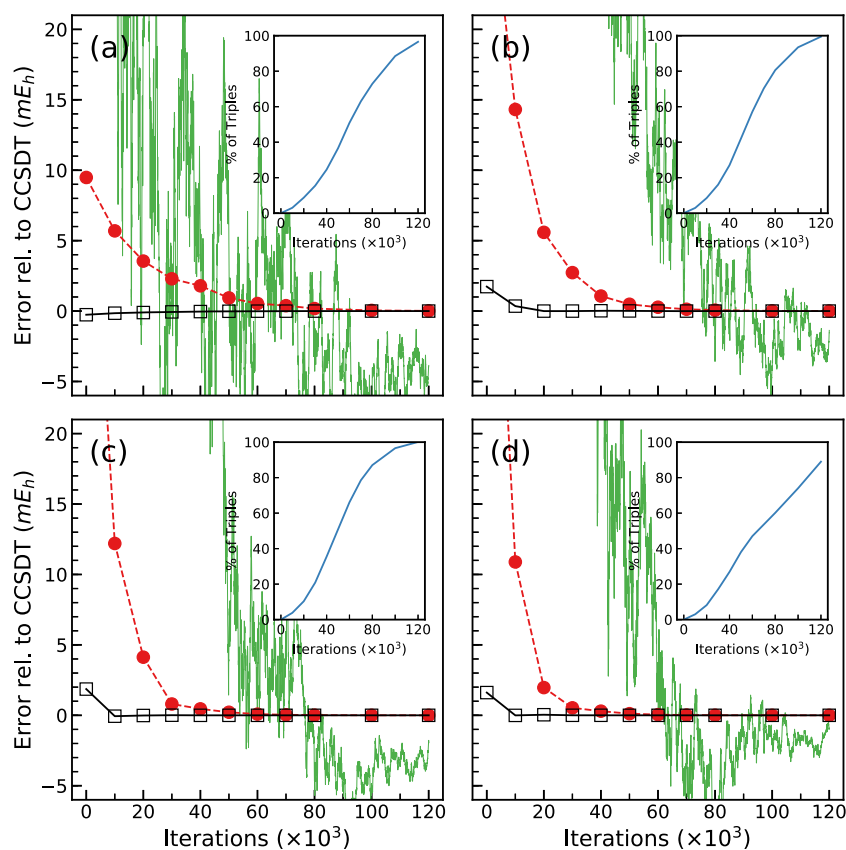
<sup>h</sup>Total CCSDT energy in hartree.

do this deterministically by turning to the previously mentioned CC( $t;3$ ) method, which uses the CC( $P;Q$ ) formalism to correct the energies obtained in the active-space CCSDt calculations for the remaining  $T_3$  correlation effects that the CCSDt approach did not capture,<sup>57,61</sup> or by the approximation to CC( $t;3$ ) that replaces the CC( $P;Q$ ) triples correction to CCSDt by its perturbative CCSD( $T$ ) analog, abbreviated as CCSD( $T$ )-h.<sup>121–123</sup> Alternatively, one can resort to the semi-stochastic CC( $P;Q$ ) framework advocated in this work, in which the same goal is accomplished by using full or truncated CIQMC propagations to identify the leading triply excited determinants for the inclusion in the underlying  $P$  space without having to use active orbitals.

The semi-stochastic CC( $P;Q$ ) results and the underlying CC( $P$ ) energies shown in Table I and Figs. 1–3 confirm the above expectations. Indeed, with only about 30%–40% of the triples in the  $P$  space, captured after the relatively short FCIQMC, CISDT-MC, and CISDTQ-MC runs at  $R = R_e$  and  $1.5R_e$  and even less than that (~15% to 20%) when the  $R = 2R_e$  and  $5R_e$  geometries are considered, the errors in the uncorrected CC( $P$ ) energies relative to their CCSDT parents are already on the order of 1 millihartree or smaller. This is a massive error reduction compared to the initial,  $\tau = 0$ , CC( $P$ ), i.e., CCSD energy values, especially at the larger F–F separations, where the differences between the CCSD and CCSDT energies are as high as 45.638 millihartree at  $R = 2R_e$  or 49.816 millihartree at  $R = 5R_e$ . The CC( $P;Q$ ) corrections based on Eq. (8) accelerate the convergence toward CCSDT even further, allowing one to reach the submillihartree accuracy levels relative to the parent CCSDT energetics almost instantaneously, out of the early stages of the FCIQMC,

CISDT-MC, and CISDTQ-MC propagations, when no more than 10% of all triples are included in the corresponding  $P$  spaces and when the total numbers of walkers used in the CIQMC runs represent tiny fractions of the walker populations required to converge these runs.

The CC( $P;Q$ )<sub>EN</sub> correction, which adopts the Epstein–Nesbet form of the  $D_K(P)$  denominator in determining the  $\ell_K(P)$  amplitudes entering Eq. (8), is particularly effective in this regard. With less than 10% triples in the stochastically determined  $P$  spaces, captured after 20 000 or fewer  $\delta\tau = 0.0001$  a.u. MC iterations, where, as shown in Figs. 1–3, the FCIQMC, CISDT-MC, and CISDTQ-MC runs are very far from convergence, the differences between the CC( $P;Q$ )<sub>EN</sub> energies and their CCSDT parents are on the order of 0.1 millihartree, being usually even smaller. This is not only true at the equilibrium geometry but also at the larger values of  $R$ , including  $R = 5R_e$ , where the F–F bond in  $F_2$  is already *de facto* broken. As shown in Table S.1 of the supplementary material, the total numbers of walkers corresponding to 20 000  $\delta\tau = 0.0001$  a.u. MC iterations initiated by placing 100 walkers on the RHF reference determinant  $|\Phi\rangle$ , which range from about 6300 to 9400 when one uses FCIQMC, 5800 to 7600 when FCIQMC is replaced by CISDTQ-MC, and 3500 to 4300 when the CISDT-MC approach is employed, represent tiny fractions of the walker populations at  $\tau = 12.0$  a.u., where we stopped our CIQMC runs (0.02%–0.53% in the case of FCIQMC, 0.07%–0.72% in the case of CISDTQ-MC, and 0.21%–1.78% in the CISDT-MC case, where total walker populations are smallest). When we perform somewhat longer FCIQMC, CISDT-MC, and CISDTQ-MC propagations, allowing them to capture about 40%–50% of the

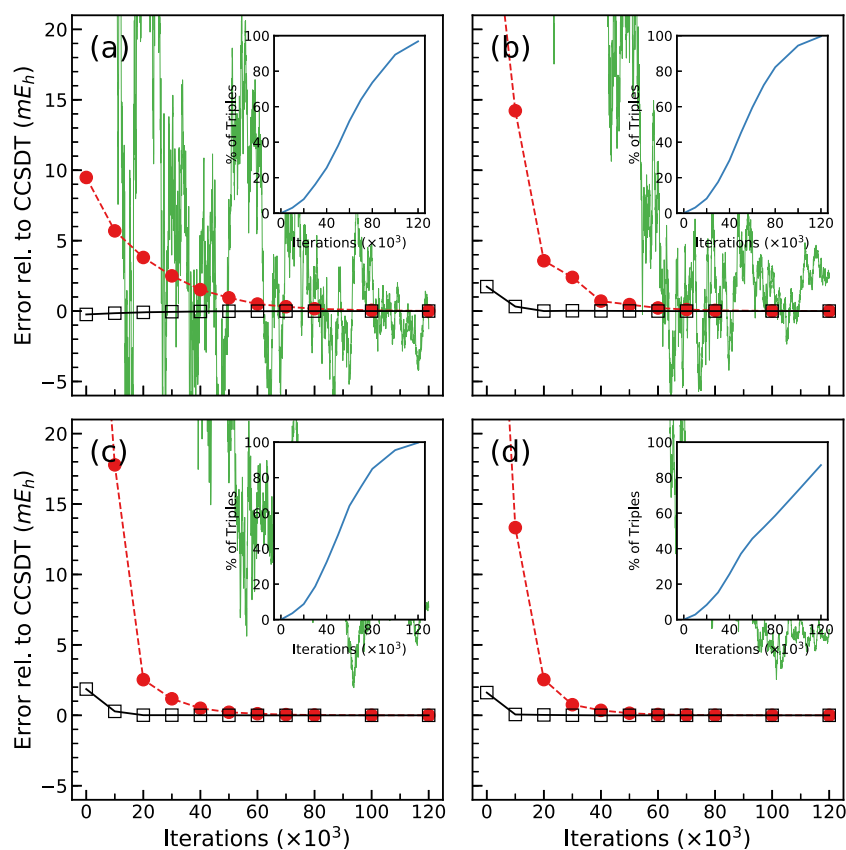


**FIG. 1.** Convergence of the  $CC(P)$  (red filled circles and dashed lines) and  $CC(P;Q)_{EN}$  (black open squares and solid lines) energies toward CCSDT for the  $F_2/cc\text{-}pVDZ$  molecule in which the F–F distance  $R$  was set at (a)  $R_e$ , (b)  $1.5R_e$ , (c)  $2R_e$ , and (d)  $5R_e$ , where  $R_e = 2.66816$  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 captured by  $i$ -FCIQMC. All energies are errors relative to CCSDT in millihartree, and the insets show the percentages of triples captured during the  $i$ -FCIQMC propagations.

triples in the  $P$  space, when  $R = R_e$  and  $1.5R_e$ , and 20%–30% when  $R \geq 2R_e$ , the  $CC(P;Q)_{EN}$  calculations recover the CCSDT energetics to within 10 or so microhartree. This happens after 50 000–60 000  $\delta\tau = 0.0001$  a.u. MC iterations, when  $R = R_e$  and  $1.5R_e$ , and 30 000–40 000 MC time steps when  $R \geq 2R_e$ , i.e., when the underlying CIQM propagations are still in their early stages (cf. Figs. 1–3). As demonstrated in Table S.1 of the [supplementary material](#), even in this case, the total numbers of walkers characterizing the FCIQMC, CISDT-MC, and CISDTQ-MC runs used to obtain these highly accurate  $CC(P;Q)_{EN}$  results remain much smaller than the walker populations required to converge the CIQMC runs. In the case of FCIQMC, they are about 60 000 or 1%–5% of the walker populations at  $\tau = 12.0$  a.u., where we stopped our CIQMC propagations, for  $R = R_e$  and  $1.5R_e$  and about 20 000–50 000 or 0.1%–0.2% of the walker populations at  $\tau = 12.0$  a.u. when the  $R \geq 2R_e$  region is explored. They are even smaller when the truncated CIQMC approaches, especially CISDT-MC, are utilized. In the case of CISDT-MC, the total numbers of walkers allowing us to converge the CCSDT energetics using the semi-stochastic  $CC(P;Q)_{EN}$  calculations to within  $\sim 10$  microhartree are as little as  $\sim 20$  000 or 5%–12% of the total walker populations at  $\tau = 12.0$  a.u., which themselves are 6–12 times smaller than those used at  $\tau = 12.0$  a.u. by FCIQMC, for  $R = R_e$  and  $1.5R_e$  and about 10 000 or 1% of the CISDT-MC walker populations at  $\tau = 12.0$  a.u., which themselves are 4%–5% of their  $\tau = 12.0$  a.u. FCIQMC counterparts, when  $R \geq 2R_e$ . The  $CC(P;Q)_{MP}$  correction,

in which the Epstein–Nesbet  $D_K(P)$  denominator, Eq. (11), in the definition of  $\ell_K(P)$  amplitudes entering Eq. (8) is replaced by its simplified Møller–Plesset form, is not as accurate as  $CC(P;Q)_{EN}$ , but it still accelerates the convergence of the underlying  $CC(P)$  energies, allowing one to recover the parent CCSDT energies to within  $\sim 0.1$  millihartree once about 40% ( $R = R_e$  and  $1.5R_e$ ) or 15%–20% ( $R = 2R_e$  and  $5R_e$ ) of the triples are captured by the FCIQMC, CISDT-MC, and CISDTQ-MC propagations.

The results shown in Table I and Figs. 1–3 demonstrate that it is practically irrelevant whether one uses FCIQMC or one of its less expensive truncated forms, such as CISDT-MC and CISDTQ-MC examined in this study, to identify the leading triply excited determinants for the inclusion in the  $P$  space used in the  $CC(P;Q)$  and the underlying  $CC(P)$  calculations. Clearly, as  $\tau$  approaches  $\infty$ , the FCIQMC, CISDT-MC, and CISDTQ-MC propagations converge to completely different limits (FCI in the case of FCIQMC, CISDT-MC in the case of CISDT-MC, and CISDTQ in the case of CISDTQ-MC), but this has virtually no impact on the convergence patterns observed in our semi-stochastic  $CC(P)$  and  $CC(P;Q)$  calculations. This is a consequence of the fact that the uncorrected  $CC(P)$  and corrected  $CC(P;Q)$  computations are capable of recovering the parent high-level CC energetics, such as those corresponding to full CCSDT discussed in this subsection, based on the information extracted from the early stages of the corresponding CIQMC runs. In particular, if we are targeting CCSDT, all we need from the CIQMC

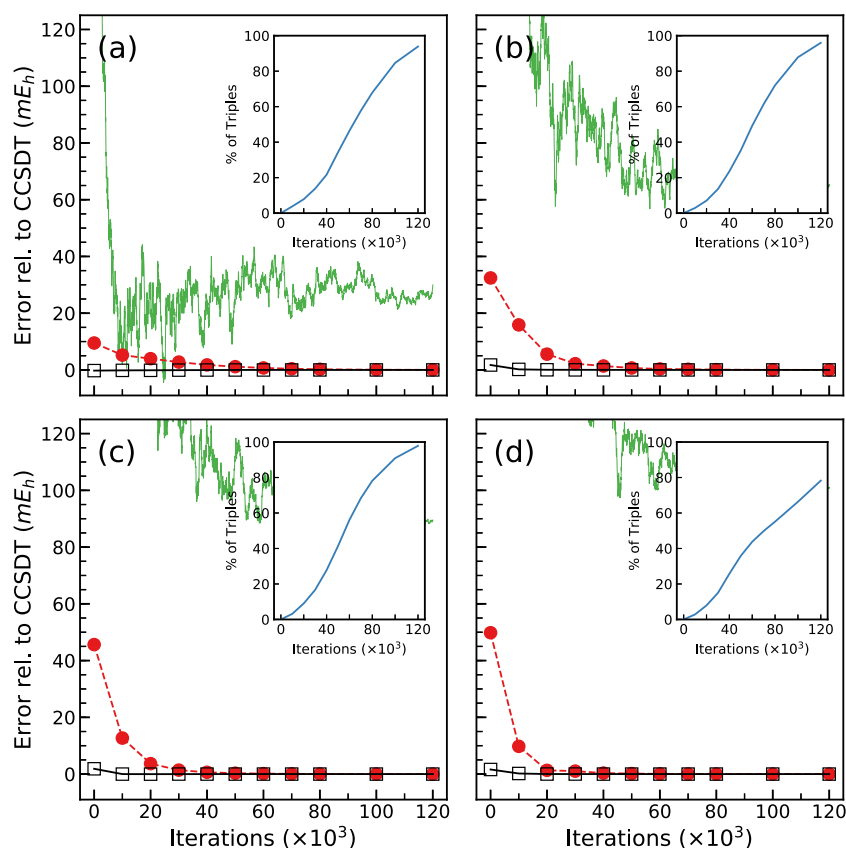


**FIG. 2.** Same as Fig. 1 except that the subsets of triples included in the  $CC(P)$  calculations are now identified by the  $i$ -CISDTQ-MC simulations and the corresponding  $Q$  spaces consist of the triples not captured by  $i$ -CISDTQ-MC. As in Fig. 1, the F-F distance  $R$  was set at (a)  $R_e$ , (b)  $1.5R_e$ , (c)  $2R_e$ , and (d)  $5R_e$ , where  $R_e = 2.668$  bohr is the equilibrium geometry.

calculations is a meaningful list of the leading triply excited determinants, which any CIQMC calculation that is allowed to sample the triples subspace of the Hilbert space, even the crude CISDT-MC approach, can provide. One can see, for example, in Table I that the fractions of triples captured by the FCIQMC, CISDT-MC, and CISDTQ-MC runs at the various numbers of MC iterations (various propagation times  $\tau$ ) are very similar. Detailed inspection of the corresponding lists of triply excited determinants shows that while the numbers of walkers on the individual determinants may substantially differ, the lists of triples identified by the FCIQMC, CISDT-MC, and CISDTQ-MC propagations, especially the more important ones that result in larger  $T_3^{(MC)}$  amplitudes in the subsequent deterministic  $CC(P)$  steps, are not much different. Once the lists of the leading triples are identified, we turn to the  $CC(P)$  computations, correcting them for the remaining triples not captured by CIQMC, and this makes the semi-stochastic  $CC(P)$  and  $CC(P;Q)$  calculations rather insensitive to the type of the CIQMC approach used to construct these lists.

All of the above observations regarding the ability of the semi-stochastic  $CC(P;Q)$  calculations using the FCIQMC, CISDT-MC, and CISDTQ-MC propagations to rapidly converge the full CCSDT energetics remain true when the cc-pVDZ basis set is replaced by its larger cc-pVTZ and aug-cc-pVTZ counterparts (both using the spherical components of  $d$  and  $f$  functions). This is illustrated in Table II, where we examine the stretched  $F_2$  molecule, in which the F-F distance  $R$  is set at  $2R_e$ . We chose  $R = 2R_e$  since, in analogy to

the previously discussed cc-pVDZ basis set, the  $T_3$  effects at this geometry, obtained by calculating differences of the respective CCSDT and CCSD energies, which are  $-62.819$  millihartree, when the cc-pVTZ basis set is employed, and  $-65.036$  millihartree, when the aug-cc-pVTZ basis set is used, are not only very large but also larger, in absolute value, than the corresponding CCSDT dissociation energies (differences between the CCSDT energies at  $R = 5R_e$ , where the F-F bond is broken, and  $R = R_e$  obtained with the cc-pVTZ and aug-cc-pVTZ basis sets are about 57 and 60 millihartree, respectively). We also chose it since the  $R = 2R_e$  stretch of the F-F bond length is large enough for the conventional CCSD(T) approach to fail in a major way when the cc-pVTZ and aug-cc-pVTZ basis sets are employed, resulting in the  $-26.354$  and  $-27.209$  millihartree errors relative to CCSDT, respectively. The CCSD(2)<sub>T</sub> correction to CCSD or the equivalent CR-CC(2,3)<sub>A</sub> approximation, represented in Table II by the  $\tau = 0$   $CC(P;Q)_{MP}$  results, helps, but large differences between the CCSD(2)<sub>T</sub> and CCSDT energies, of 9.211 millihartree in the cc-pVTZ case and 9.808 millihartree when the aug-cc-pVTZ basis set is employed, remain. The CR-CC(2,3)<sub>D</sub> approach, represented in Table II by the  $\tau = 0$   $CC(P;Q)_{EN}$  data, is more effective than other triples corrections to CCSD, reducing the large errors relative to CCSDT observed in the CCSD(T) and CCSD(2)<sub>T</sub> calculations to 4.254 (cc-pVTZ) and 5.595 (aug-cc-pVTZ) millihartree, but none of the above results are as good as the energies resulting from the semi-stochastic  $CC(P;Q)$  calculations using FCIQMC, CISDT-MC, and CISDTQ-MC.



**FIG. 3.** Same as Fig. 1 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 captured by *i*-CISDT-MC. As in Fig. 1, 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.

Indeed, as shown in Table II, we observe a rapid error reduction relative to the parent CCSDT data once we start migrating the triply excited determinants identified during the FCIQMC, CISDT-MC, and CISDTQ-MC propagations into the underlying *P* space. With about 20%–30% (cc-pVTZ) or 30%–40% (aug-cc-pVTZ) of the triples in the *P* space, the 62.819 and 65.036 millihartree errors resulting from the initial CCSD [ $\tau = 0$  CC(*P*)] computations decrease to a 1–2 millihartree level when the CC(*P*) method is employed. The CC(*P*; *Q*) corrections due to the remaining triples not captured by FCIQMC, CISDT-MC, and CISDTQ-MC accelerate the convergence toward CCSDT even further, with the semistochastic CC(*P*; *Q*)<sub>EN</sub> approach being particularly efficient in this regard. With only 2%–4% of the triples in the stochastically determined *P* spaces, captured after 20 000–30 000  $\delta\tau = 0.0001$  a.u. MC iterations, which are the very early stages of the FCIQMC, CISDT-MC, and CISDTQ-MC propagations, the CC(*P*; *Q*)<sub>EN</sub> calculations recover the full CCSDT energetics corresponding to the cc-pVTZ and aug-cc-pVTZ basis sets to within 0.1–0.2 millihartree. After 50 000 (cc-pVTZ) or 60 000 (aug-cc-pVTZ) MC iterations, where the FCIQMC, CISDT-MC, and CISDTQ-MC runs are still far from convergence, capturing only about 20%–30% (cc-pVTZ) or 30%–40% (aug-cc-pVTZ) of the triples, the errors in the CC(*P*; *Q*)<sub>EN</sub> energies relative to CCSDT reduce to a 10 microhartree level. Similar to the previously discussed calculations using the cc-pVDZ basis set, the total numbers of walkers characterizing the CIQMC

propagations that allowed us to reproduce the CCSDT/cc-pVTZ and CCSDT/aug-cc-pVTZ energies so accurately represent tiny fractions of the walker populations required to converge the CIQMC runs (see Table S.2 of the supplementary material). For example, the total number of walkers corresponding to 30 000  $\delta\tau = 0.0001$  a.u. FCIQMC iterations initiated by placing 100 walkers on the RHF reference determinant, which enable the FCIQMC-driven CC(*P*; *Q*)<sub>EN</sub> approach to recover the CCSDT/aug-cc-pVTZ energy of  $F_2$  at  $R = 2R_e$  to within  $\sim 0.1$  millihartree, is only about 200 000. As shown in Table S.2 of the supplementary material, this translates into less than 0.1% of the total walker population used by the FCIQMC run at  $\tau = 10.0$  a.u., where we terminated our CIQMC propagations. With about  $2 \times 10^6$  walkers in the FCIQMC computation, reached after 50 000  $\delta\tau = 0.0001$  a.u. MC iterations, i.e., with less than 1% of the total walker population at  $\tau = 10.0$  a.u., the difference between the FCIQMC-based CC(*P*; *Q*)<sub>EN</sub> energy and its CCSDT parent reduces to 16 microhartree. The analogous  $\tau = 5.0$  a.u. CISDTQ-MC and CISDT-MC calculations, which allow the CC(*P*; *Q*)<sub>EN</sub> approach to recover the CCSDT/aug-cc-pVTZ energy of  $F_2$  at  $R = 2R_e$  to within 19 and 38 microhartree, respectively, use even smaller numbers of walkers, namely, a little over  $1 \times 10^6$  in the case of CISDTQ-MC and less than 300 000 in the CISDT-MC case. In analogy to the cc-pVDZ basis set, the CC(*P*; *Q*)<sub>MP</sub> correction is less accurate than its CC(*P*; *Q*)<sub>EN</sub> counterpart when the cc-pVTZ and aug-cc-pVTZ basis sets are employed, recovering the CCSDT energetics to within

**TABLE II.** 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			$CC(P)^a$			$CC(P;Q)_{MP}^a$			$CC(P;Q)_{EN}^a$		
		FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>
VTZ	0		0			62.819 <sup>e</sup>			9.211 <sup>f</sup>			4.254 <sup>g</sup>	
	10 000	1	1	1	29.714	31.973	31.571	2.738	3.104	2.636	0.728	0.896	0.539
	20 000	2	2	2	11.179	14.687	20.194	0.824	1.097	1.487	0.071	0.151	0.217
	30 000	6	6	4	5.787	6.031	9.294	0.400	0.425	0.617	0.028	0.030	0.025
	40 000	14	14	10	2.406	2.574	4.203	0.160	0.171	0.284	0.002	0.001	0.014
	50 000	27	26	19	1.193	1.237	2.177	0.076	0.078	0.138	−0.003	−0.002	−0.002
	60 000	42	42	30	0.490	0.489	1.144	0.029	0.029	0.071	−0.002	−0.002	−0.005
	70 000	59	57	44	0.178	0.171	0.576	0.011	0.010	0.037	−0.001	−0.001	−0.002
	80 000	72	71	56	0.045	0.054	0.309	0.003	0.003	0.020	0.000	0.000	−0.001
100 000	90	89	78	0.002	0.003	0.130	0.000	0.000	0.009	0.000	0.000	0.000	
∞		100			−199.238 344 <sup>h</sup>			...			...		
AVTZ	0		0			65.036 <sup>e</sup>			9.808 <sup>f</sup>			5.595 <sup>g</sup>	
	10 000	0	0	0	36.316	38.874	42.801	3.641	4.144	4.851	1.594	1.786	2.304
	20 000	1	1	1	17.190	20.799	26.557	1.276	1.656	2.288	0.382	0.512	0.791
	30 000	4	4	3	8.065	9.272	13.279	0.549	0.623	0.928	0.138	0.138	0.246
	40 000	10	10	7	4.408	4.677	7.477	0.291	0.307	0.499	0.057	0.062	0.106
	50 000	23	22	15	2.208	2.425	3.951	0.136	0.150	0.244	0.016	0.019	0.038
	60 000	41	39	27	1.021	1.137	2.052	0.058	0.070	0.124	0.002	0.005	0.013
	70 000	61	58	61	0.385	0.455	0.385	0.021	0.025	0.059	0.000	0.000	0.001
	80 000	78	76	78	0.125	0.154	0.125	0.007	0.008	0.026	0.000	0.000	0.000
	100 000	97	96	97	0.007	0.009	0.007	0.000	0.001	0.004	0.000	0.000	0.000
∞		100			−199.253 022 <sup>h</sup>			...			...		

<sup>a</sup>Unless otherwise stated, all energies are reported as errors relative to CCSDT in millihartree.

<sup>b</sup>FCI stands for  $i$ -FCIQMC.

<sup>c</sup>CIQ stands for  $i$ -CISDTQ-MC.

<sup>d</sup>CIT stands for  $i$ -CISDT-MC.

<sup>e</sup>Equivalent to CCSD.

<sup>f</sup>Equivalent to the CCSD energy corrected for the effects of  $T_3$  clusters using the CCSD(2)<sub>T</sub> approach of Ref. 88, which is equivalent to the approximate form of the completely renormalized CR-CC(2,3) approach of Refs. 75 and 76, abbreviated sometimes as CR-CC(2,3)<sub>A</sub> or CR-CC(2,3)<sub>A</sub>.<sup>62,63,78,80,83</sup>

<sup>g</sup>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. 75 and 76, abbreviated sometimes as CR-CC(2,3)<sub>D</sub> or CR-CC(2,3)<sub>D</sub>.<sup>62,63,78,80,83</sup>

<sup>h</sup>Total CCSDT energy in hartree.

0.1–0.2 millihartree after 50 000 rather than 20 000–30 000 MC iterations, i.e., after about 20%–30% rather than 2%–4% of the triples are captured by the CIQMC propagations, but the overall error reduction compared to the underlying  $CC(P)$  calculations or the various noniterative triples corrections to CCSD is still impressive.

Similar to the cc-pVDZ basis set, the semi-stochastic  $CC(P;Q)$  calculations using larger cc-pVTZ and aug-cc-pVTZ bases are rather insensitive to the type of the CIQMC approach used to identify the leading triples for the inclusion in the  $P$  space. Based on the results in Table II, one might try to argue that the energies obtained with the uncorrected  $CC(P)$  approach using the CISDT-MC propagations are characterized by slower convergence compared to their CISDTQ-MC- and FCIQMC-driven counterparts, but this would

be misleading, since CISDT-MC captures the leading triples at a somewhat slower rate, while being less expensive than CISDTQ-MC and FCIQMC at the same time. For example, the CISDT-MC-driven  $CC(P)$  computations for  $F_2$  at  $R = 2R_e$  using the cc-pVTZ basis set need 60 000  $\delta\tau = 0.0001$  a.u. MC iterations to reach a  $\sim 1$  millihartree accuracy relative to the corresponding CCSDT energy. The  $CC(P)$  approach using CISDTQ-MC and FCIQMC reaches the same accuracy level sooner, after 50 000 MC iterations. One should keep in mind, however, that it takes 60 000 MC time steps for the CISDT-MC propagation to capture about 30% of the triples, needed to reach a  $\sim 1$  millihartree accuracy level in the subsequent  $CC(P)$  calculations, and the analogous CISDTQ-MC and FCIQMC runs capture a similar fraction of the triples after 50 000 time steps. Ultimately,

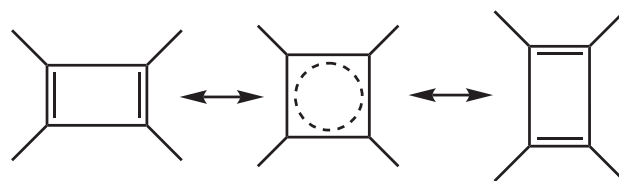
one needs to remember that all CIQMC-driven  $CC(P)$  computations considered in this subsection converge to CCSDT as  $\tau \rightarrow \infty$ , independent of the type of the CIQMC approach used to define the underlying  $P$  spaces, as long as the CIQMC propagation is allowed to spawn walkers on the triply excited determinants. Perhaps more importantly, the  $CC(P;Q)$  corrections to the  $CC(P)$  energies make the convergence toward CCSDT not only much faster but also less dependent on the type of the CIQMC approach used in the calculations since they take care of the triples that were not captured by the respective QMC propagations.

Before discussing our next molecular example, it is worth pointing out that the FCIQMC-driven  $CC(P;Q)$  calculations reported in Tables I and II and Fig. 1, in which, as explained in Sec. II B, we used complete representations of  $\bar{H}^{(P)}$  and  $\Lambda^{(P)}$  in determining corrections  $\delta(P;Q)$ , approach the parent CCSDT energetics of the stretched  $F_2$  system in the early stages of the underlying FCIQMC propagations faster than the analogous calculations reported in Ref. 58, where the three-body component of  $\Lambda^{(P)}$  was neglected. For example, the  $CC(P;Q)$  energies of  $F_2$  at  $R = 2R_e$  using the aug-cc-pVTZ basis set obtained in this work after 10 000, 20 000, and 30 000  $\delta\tau = 0.0001$  a.u. MC iterations of the underlying FCIQMC propagation differ from the corresponding CCSDT energy by 1.594, 0.382, and 0.138 millihartree, respectively (see Table II). The analogous energy differences reported in Ref. 58, of 3.770, 1.661, and 0.454 millihartree, respectively, are noticeably larger (see Table II in the supplementary material of Ref. 58). In fact, by comparing the FCIQMC-, CISDT-MC-, and CISDTQ-MC-based  $CC(P;Q)$  energies shown in Tables I and II and Figs. 1–3, determined by treating the de-excitation operator  $\Lambda^{(P)}$  in Eq. (9) fully, i.e., by defining  $\Lambda^{(P)}$  as  $\Lambda_1 + \Lambda_2 + \Lambda_3^{(MC)}$ , with their FCIQMC- and CCSDT-MC-based counterparts obtained in Ref. 58, where  $\Lambda^{(P)}$  was approximated by  $\Lambda_1 + \Lambda_2$ , we can conclude that as long as  $\Lambda_3^{(MC)}$  is not neglected, one can replace FCIQMC by CISDTQ-MC or, even, CISDT-MC and still improve the rate of convergence of the  $CC(P;Q)$  energies toward CCSDT in the early stages of the QMC propagations compared to that reported in Ref. 58.

The above observations, combined with the superior performance of the  $CC(P;Q)_{EN}$  approach compared to its  $CC(P;Q)_{MP}$  counterpart, suggest that a complete treatment of correction  $\delta(P;Q)$ , as dictated by Eqs. (8), (9), and (11), is more important, especially when one is interested in accelerating convergence of the semi-stochastic  $CC(P;Q)$  calculations for stretched or more multireference molecules in the early stages of the QMC propagations, than the actual type of the underlying CIQMC approach. It is interesting to examine if the same remains true when other molecular examples, including those discussed in Secs. III B and III C, are considered.

## B. Automerization of cyclobutadiene

Our next example is the challenging and frequently studied<sup>61,116,117,124–141</sup> automerization of cyclobutadiene (see Fig. 4). In this case, in order to obtain reliable energetics using computational means, especially the activation energy, one has to provide an accurate and well-balanced description of the nondegenerate closed-shell reactant (or the equivalent product) species, in which the many-electron correlation effects have a predominantly dynamical character, and the quasi-degenerate, biradicaloid transition state



**FIG. 4.** The key molecular structures defining the automerization of cyclobutadiene. The leftmost and rightmost structures represent the degenerate reactant/product minima, whereas the structure in the center corresponds to the transition state.

characterized by substantial non-dynamical correlations. Experiment suggests that the activation energy for the automerization of cyclobutadiene is somewhere between 1.6 and 10 kcal/mol.<sup>124,126</sup> The most accurate single- and multi-reference calculations performed to date, reviewed, for example, in Refs. 61, 117, and 140, imply that the purely electronic value of the energy barrier falls into the 6–10 kcal/mol range. In particular, as pointed out in Ref. 61 (see also Ref. 116), one can obtain a reliable description of the activation energy using the full CCSDT approach. Given this information and the methodological nature of the present study, in which we had to perform a large number of semi-stochastic  $CC(P)$  and  $CC(P;Q)$  calculations, exploring three different types of the CIQMC method, including FCIQMC, CISDT-MC, and CISDTQ-MC, and probing many values of the QMC propagation time  $\tau$ , in a discussion below, we focus on converging the CCSDT energetics obtained using the spherical cc-pVDZ basis set. As shown in Ref. 61 and Table III, the CCSDT/cc-pVDZ activation energy characterizing the automerization of cyclobutadiene, assuming the reactant/product and transition-state geometries obtained with the multireference average-quadratic CC (MR-AQCC) approach<sup>142,143</sup> in Ref. 134, which we adopt in the  $CC(P)$  and  $CC(P;Q)$  calculations reported in this work as well, is 7.627 kcal/mol, in reasonable agreement with the most accurate *ab initio* results reported to date. The results of our semi-stochastic  $CC(P)$  and  $CC(P;Q)$  calculations, aimed at recovering the CCSDT/cc-pVDZ energetics of the reactant and transition-state species and the corresponding activation energy using the FCIQMC, CISDT-MC, and CISDTQ-MC propagations to identify the leading triply excited determinants for constructing the underlying  $P$  spaces, are summarized in Table III and Fig. 5.

As already mentioned, all of the noniterative triples corrections to CCSD, including CCSD(T),  $\Lambda$ -CCSD(T), CCSD(2)<sub>T</sub>, and CR-CC(2,3), perform very poorly in this case, producing activation barriers in a 16–17 kcal/mol range when the cc-pVDZ basis set is considered,<sup>61,117</sup> instead of ~8 kcal/mol obtained with CCSDT (it should be noted that the 16–17 kcal/mol values are also way outside the experimentally derived and most accurate theoretically determined ranges of 1.6–10 kcal/mol and 6–10 kcal/mol, respectively). They improve the CCSD activation energy, which is even worse [about 21 kcal/mol; see the  $\tau = 0$   $CC(P)$  barrier in Table III], but the improvements offered by the noniterative triples corrections to CCSD are far from sufficient. This, in particular, applies to the CCSD(2)<sub>T</sub> = CR-CC(2,3)<sub>A</sub> and CR-CC(2,3)<sub>D</sub> approaches, represented in Table III by the  $\tau = 0$   $CC(P;Q)_{MP}$  and  $CC(P;Q)_{EN}$  data, respectively, where errors in the resulting activation energies relative to CCSDT are 9.611 kcal/mol (126%) in the former case and

**TABLE III.** 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 defining the automerization of cyclobutadiene, as described by the cc-pVDZ basis set, optimized in the MR-AQCC calculations reported in Ref. 134, 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 iterations	% of triples			$CC(P)^a$			$CC(P;3)_{MP}^a$			$CC(P;3)_{EN}^a$		
		FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>	FCI <sup>b</sup>	CIQ <sup>c</sup>	CIT <sup>d</sup>
R	0		0			26.827 <sup>e</sup>		4.764 <sup>f</sup>				0.848 <sup>g</sup>	
	10 000	0	0	0	25.758	25.985	25.484	4.437	4.535	4.324	0.696	0.763	0.625
	20 000	2	2	1	22.532	22.513	22.462	3.684	3.621	3.612	0.496	0.418	0.433
	30 000	6	5	5	17.369	17.857	18.880	2.599	2.676	2.889	0.230	0.228	0.279
	40 000	16	15	12	11.845	12.034	13.834	1.635	1.649	2.007	0.092	0.080	0.164
	50 000	31	30	24	6.895	7.176	9.202	0.877	0.913	1.235	0.022	0.023	0.057
	60 000	52	51	41	3.273	3.524	5.205	0.386	0.417	0.645	0.001	0.000	0.010
	70 000	72	70	59	1.321	1.498	2.594	0.146	0.170	0.302	-0.003	-0.002	-0.003
	80 000	85	84	75	0.512	0.563	1.181	0.056	0.060	0.131	-0.001	-0.001	-0.002
	$\infty$		100			-154.244	157 <sup>h</sup>		...			...	
TS	0		0			47.979 <sup>e</sup>		20.080 <sup>f</sup>				14.636 <sup>g</sup>	
	10 000	0	0	0	45.875	46.427	45.777	18.899	19.135	18.037	13.680	13.842	12.665
	20 000	1	2	1	39.577	37.689	39.655	14.220	12.522	13.774	9.452	7.793	8.863
	30 000	5	5	5	30.836	28.405	33.111	9.660	7.404	10.798	5.785	3.648	6.651
	40 000	15	13	13	18.976	19.811	23.797	4.046	4.313	6.457	1.556	1.661	3.367
	50 000	31	27	26	9.795	9.727	12.495	1.634	1.488	2.238	0.309	0.243	0.602
	60 000	52	47	42	3.936	4.136	6.217	0.501	0.525	0.886	0.026	0.025	0.105
	70 000	70	67	60	1.491	1.488	2.841	0.173	0.168	0.363	0.003	0.001	0.018
	80 000	84	82	74	0.525	0.591	1.260	0.058	0.065	0.148	0.000	0.000	0.001
	$\infty$		100			-154.232	002 <sup>h</sup>		...			...	
Barrier	0		0/0			13.274 <sup>e</sup>		9.611 <sup>f</sup>				8.653 <sup>g</sup>	
	10 000	0/0	0/0	0/0	12.624	12.828	12.734	9.075	9.162	8.605	8.148	8.208	7.555
	20 000	2/1	2/2	1/1	10.696	9.523	10.789	6.612	5.586	6.377	5.620	4.628	5.290
	30 000	6/5	5/5	5/5	8.450	6.619	8.931	4.431	2.967	4.963	3.487	2.146	3.999
	40 000	16/15	15/13	12/13	4.475	4.881	6.252	1.513	1.672	2.793	0.919	0.992	2.011
	50 000	31/31	30/27	24/26	1.820	1.601	2.067	0.475	0.361	0.629	0.181	0.138	0.343
	60 000	52/52	51/47	41/42	0.416	0.384	0.635	0.073	0.068	0.151	0.016	0.016	0.060
	70 000	72/70	70/67	59/60	0.107	-0.006	0.155	0.017	-0.001	0.038	0.003	0.002	0.013
	80 000	85/84	84/82	75/74	0.008	0.018	0.050	0.001	0.003	0.011	0.001	0.001	0.002
	$\infty$		100/100			7.627 <sup>i</sup>			...			...	

<sup>a</sup>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.

<sup>b</sup>FCI stands for  $i$ -FCIQMC.

<sup>c</sup>CIQ stands for  $i$ -CISDTQ-MC.

<sup>d</sup>CIT stands for  $i$ -CISDT-MC.

<sup>e</sup>Equivalent to CCSD.

<sup>f</sup>Equivalent to the CCSD energy corrected for the effects of  $T_3$  clusters using the CCSD(2)<sub>T</sub> approach of Ref. 88, which is equivalent to the approximate form of the completely renormalized CR-CC(2,3) approach of Refs. 75 and 76, abbreviated sometimes as CR-CC(2,3)<sub>A</sub> or CR-CC(2,3)<sub>A</sub>.<sup>62,63,78,80,83</sup>

<sup>g</sup>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. 75 and 76, abbreviated sometimes as CR-CC(2,3)<sub>D</sub> or CR-CC(2,3)<sub>D</sub>.<sup>62,63,78,80,83</sup>

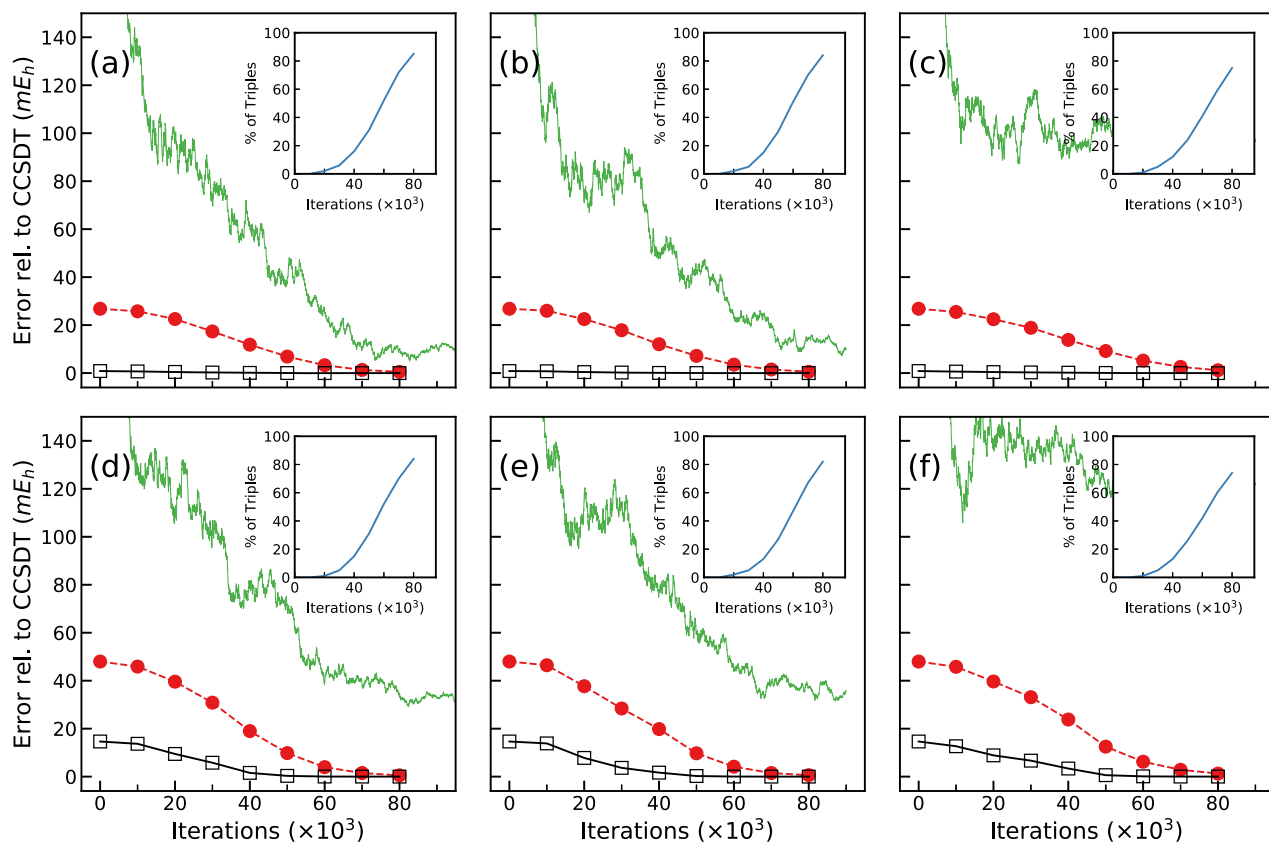
<sup>h</sup>Total CCSDT energy in hartree.

<sup>i</sup>The CCSDT activation barrier in kcal/mol.

8.653 kcal/mol (113%) in the case of the latter method. As explained in Ref. 61, the poor performance of the noniterative triples corrections to CCSD in describing the automerization of cyclobutadiene is a consequence of neglecting the coupling between the  $T_3$  clusters and their lower-order  $T_1$  and  $T_2$  counterparts, which is

accounted for in CCSDT but ignored in methods such as CCSD(T),  $\Lambda$ -CCSD(T), CCSD(2)<sub>T</sub>, and CR-CC(2,3). This coupling is particularly large at the transition-state geometry, where the magnitude of  $T_3$  contributions, defined as the absolute value of the difference between the CCSDT and CCSD energies, is nearly 48





**FIG. 5.** 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 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. Panels (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 captured 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 captured 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 captured by  $i$ -CISDT-MC. All reported energies are errors relative to CCSDT in millihartree. The insets show the percentages of triples captured during the relevant  $i$ -CIQMC propagations.

millihartree, when the cc-pVDZ basis set is employed, and where errors in the CCSD(T),  $\Lambda$ -CCSD(T), CCSD(2)<sub>T</sub>, and CR-CC(2,3) energies relative to CCSDT range from about 14 to 20 millihartree, as opposed to  $\sim 1$  to 5 millihartree obtained for the reactant (see Refs. 61 and 117 and Table III). In analogy to bond breaking in  $F_2$ , if we want to capture the coupling of the  $T_1$ ,  $T_2$ , and  $T_3$  clusters without having to solve full CCSDT equations while preserving the idea of noniterative triples corrections to energies obtained in lower-order CC calculations, we must solve for the  $T_1$  and  $T_2$  amplitudes, which in the CCSD(T),  $\Lambda$ -CCSD(T), CCSD(2)<sub>T</sub>, and CR-CC(2,3) approaches are obtained with CCSD, in the presence of the dominant  $T_3$  components by incorporating some triples in the iterative CC steps, and then correct the resulting energies for the remaining  $T_3$  effects neglected in the CC iterations. Again, this can be done

deterministically by solving the active-space CCSDt equations, in which the dominant  $T_3$  amplitudes are selected using active orbitals, and correcting the CCSDt energies for the remaining  $T_3$  correlations using the  $CC(P;Q)$  corrections  $\delta(P;Q)$ , as in the  $CC(t;3)$  calculations reported in Ref. 61, or by turning to the semi-stochastic form of the  $CC(P;Q)$  formalism pursued in this study, which eliminates the need for defining active orbitals, when identifying the leading triples, by resorting to CIQMC propagations. Interestingly, using the CCSD(T)-type correction to CCSDt, as in the aforementioned CCSD(T)-h approach, in the calculations for the automerization of cyclobutadiene worsens the activation energies obtained with CCSDt, moving them away from their parent CCSDT values.<sup>61</sup> This underlines the significance of treating corrections  $\delta(P;Q)$  due to the correlation effects outside the underlying  $P$  spaces as

completely as possible, following Eqs. (8), (9), and (11), avoiding drastic approximations in these equations that lead to the triples corrections of CCSD(T).

As shown in Table III and Fig. 5, the semi-stochastic CC( $P;Q$ ) calculations using FCIQMC, CISDT-MC, and CISDTQ-MC are remarkably efficient in capturing the desired  $T_3$  correlation effects. Independent of the type of the CIQMC approach, they allow us to converge the CCSDT values of the transition-state and activation energies, which are poorly described by the noniterative triples corrections to CCSD, to within 1–2 millihartree or 1–2 kcal/mol out of the early stages of the CIQMC propagations while further improving an accurate description of the reactant by methods such as CR-CC(2,3)<sub>D</sub>. Similar to F<sub>2</sub>, the performance of the CC( $P;Q$ )<sub>EN</sub> approach, which uses the Epstein–Nesbet form of the  $D_K(P)$  denominator in calculating the  $\ell_K(P)$  amplitudes entering Eq. (8), is particularly impressive. With just 5%–6% of the triples in the stochastically determined  $P$  spaces, captured by the FCIQMC, CISDT-MC, and CISDTQ-MC propagations after 30 000  $\delta\tau = 0.0001$  a.u. MC iterations, i.e., almost instantaneously, the CC( $P;Q$ )<sub>EN</sub> approach reduces the initial 0.848 millihartree, 14.636 millihartree, and 8.653 kcal/mol errors in the reactant, transition-state, and activation energies relative to CCSDT obtained in the  $\tau = 0$  CC( $P;Q$ )<sub>EN</sub> or CR-CC(2,3)<sub>D</sub> calculations by factors of 2–4, to 0.228–0.279 millihartree, 3.648–6.651 millihartree, and 2.146–3.999 kcal/mol, respectively. After the additional 10 000 MC time steps, which result in capturing 12%–16% of the triples in the underlying  $P$  spaces, errors in the CC( $P;Q$ )<sub>EN</sub> reactant, transition-state, and activation energies relative to their CCSDT values become 0.080–0.164 millihartree, 1.556–3.367 millihartree, and 0.919–2.011 kcal/mol, respectively. These are remarkable improvements compared to the initial CR-CC(2,3)<sub>D</sub> values, especially if we realize that the early stages of the FCIQMC, CISDT-MC, and CISDTQ-MC calculations, such as 30 000–40 000  $\delta\tau = 0.0001$  a.u. MC time steps, are all very fast, using, as shown in Table S.3 of the supplementary material, tiny fractions of the total walker populations at  $\tau = 8.0$  a.u., where we terminated our CIQMC runs, and 5%–6% or 12%–16% are small fractions of the triples that result in large speedups in the underlying CC( $P$ ) calculations and significant reductions in the  $T_3$  amplitude storage requirements. After 50 000 MC iterations, where, as shown in Fig. 5, the FCIQMC, CISDT-MC, and CISDTQ-MC runs are still far from convergence, capturing about 20%–30% of the triples, i.e., still relatively small fractions of all triply excited determinants, the CC( $P;Q$ )<sub>EN</sub> calculations recover the CCSDT values of the reactant, transition-state, and activation energies to within 22–57 microhartree, 0.243–0.602 millihartree, and 0.138–0.343 kcal/mol, respectively, which is a massive error reduction compared to CR-CC(2,3)<sub>D</sub> and other noniterative triples corrections to CCSD. Again, as demonstrated in Table S.3 of the supplementary material, the total numbers of walkers used by the underlying CIQMC calculations, which allowed the semi-stochastic CC( $P;Q$ )<sub>EN</sub> computations to converge the CCSDT energetics so tightly, are not only small fractions of the corresponding walker populations at  $\tau = 8.0$  a.u., where we stopped our CIQMC propagations (about 5% in the case of FCIQMC, 8%–9% in the CISDTQ-MC case, and 15%–16% when the CISDT-MC approach was employed), but also small in absolute values. In the case of the  $\tau = 5.0$  a.u. FCIQMC and CISDTQ-MC computations corresponding to 50 000  $\delta\tau = 0.0001$  a.u. MC time steps, they are about  $2.3 \times 10^6$  and  $1.8 \times 10^6$ – $2.0 \times 10^6$ , respectively. When one switches to

CISDT-MC, they go down to less than half a million. As in the case of bond breaking in F<sub>2</sub>, the CC( $P;Q$ )<sub>MP</sub> correction, which uses the Møller–Plesset  $D_K(P)$  denominator in Eq. (9) instead of its more elaborate Epstein–Nesbet form given by Eq. (11), is less accurate than its CC( $P;Q$ )<sub>EN</sub> counterpart, but its ability to accelerate convergence of the underlying CC( $P$ ) energies and improving the results obtained with CR-CC(2,3) and other triples corrections to CCSD is still quite impressive. For example, with about 20%–30% of the triples captured by the FCIQMC, CISDT-MC, and CISDTQ-MC propagations after 50 000 MC iterations, the differences between the CC( $P;Q$ )<sub>MP</sub> reactant, transition-state, and activation energies and their CCSDT counterparts, of 0.877–1.235 millihartree, 1.488–2.238 millihartree, and 0.361–0.629 kcal/mol, are much smaller than the analogous errors relative to CCSDT resulting from the corresponding CC( $P$ ) calculations, which are 6.895–9.202 millihartree, 9.727–12.495 millihartree, and 1.601–2.067 kcal/mol, respectively, although they are not as small as the aforementioned 22–57 microhartree, 0.243–0.602 millihartree, and 0.138–0.343 kcal/mol errors obtained using the CC( $P;Q$ )<sub>EN</sub> correction.

In analogy to the fluorine molecule, the semi-stochastic CC( $P;Q$ ) calculations aimed at converging the CCSDT results for the automerization of cyclobutadiene are generally insensitive to the type of the CIQMC approach used to identify the leading triples for the inclusion in the underlying  $P$  spaces. It is sufficient to resort to the least expensive forms of the CIQMC propagations capable of capturing the triples, such as CISDT-MC or CISDTQ-MC, to obtain the fast convergence of the CC( $P;Q$ ) reactant, transition-state, and activation energies toward their CCSDT parents observed in Table III and Fig. 5. Treating the CC( $P;Q$ ) correction  $\delta(P;Q)$  fully, following Eqs. (8), (9), and (11), is, however, important. We have already discussed the benefits of using the Epstein–Nesbet form of the  $D_K(P)$  denominator, Eq. (11), in determining the  $\ell_K(P)$  amplitudes entering Eq. (8). A complete treatment of the de-excitation operator  $\Lambda^{(P)}$  in Eq. (9), which in the case of the triples corrections to the CC( $P$ ) energies considered here means representing it as  $\Lambda_1 + \Lambda_2 + \Lambda_3^{(MC)}$ , is important too. One can consider an approximation in which the three-body component  $\Lambda_3^{(MC)}$  is neglected, which is what we did in Ref. 58, but it is generally better, especially in the earlier stages of the CIQMC propagations, to keep all of the relevant many-body components of  $\Lambda^{(P)}$  in calculating the  $\ell_K(P)$  amplitudes that enter the CC( $P;Q$ ) correction  $\delta(P;Q)$ . This can be illustrated by comparing the results of the FCIQMC-driven CC( $P;Q$ ) computations shown in Table III, where we used a complete representation of  $\Lambda^{(P)}$ , in which the three-body component  $\Lambda_3^{(MC)}$  was included, with the analogous results reported in Ref. 58, where  $\Lambda_3^{(MC)}$  was neglected. For example, the differences between the CC( $P;Q$ )<sub>EN</sub> reactant, transition-state, and activation energies and their CCSDT counterparts obtained in this work after 40 000  $\delta\tau = 0.0001$  a.u. time steps of the FCIQMC propagation are 92 microhartree, 1.556 millihartree, and 0.919 kcal/mol, respectively. The analogous energy differences reported in Ref. 58 are 0.489 millihartree, 3.235 millihartree, and 1.7 kcal/mol, respectively, i.e., they are substantially larger. Ultimately, when the propagation time  $\tau$  becomes longer, different ways of handling the  $\Lambda^{(P)}$  operator or different ways of defining the  $D_K(P)$  denominator in Eq. (9) become less important, but if we are interested in accurately approximating the parent CC energetics in the

early stages of the underlying CIQMC propagations, treating these quantities fully is essential.

As shown in this subsection and Sec. III A, using complete representations of the  $\Lambda^{(P)}$  and  $\bar{H}^{(P)}$  operators and the Epstein–Nesbet-type denominators  $D_K(P)$  in determining corrections  $\delta(P;Q)$  benefits the semi-stochastic  $CC(P;Q)$  calculations aimed at converging the CCSDT energetics. In Sec. III C, which is the final part of our discussion of the numerical results obtained in this work, we investigate if similar applies to the CIQMC-driven  $CC(P;Q)$  computations targeting CCSDTQ.

### C. Double dissociation of H<sub>2</sub>O

Our last example, which illustrates the ability of the semi-stochastic  $CC(P)$  and  $CC(P;Q)$  approaches to converge the CCSDTQ energetics, is the  $C_{2v}$ -symmetric cut of the ground-state potential energy surface of the water molecule, in which both O–H bonds are simultaneously stretched without changing the  $\angle$  (H–O–H) angle, resulting in large  $T_3$  and  $T_4$  contributions. Following Ref. 118 and consistent with our earlier deterministic  $CC(P;Q)$  study,<sup>63</sup> where we also obtained the reference CCSDTQ energies, we used the spherical cc-pVDZ basis set, correlated all electrons, and considered four stretches of the O–H bonds, including  $R_{O-H} = 1.5R_e, 2R_e, 2.5R_e,$  and  $3R_e$ , in addition to the equilibrium geometry,  $R_{O-H} = R_e$ . We used the same geometries, which the reader can find in Ref. 118, in the semi-stochastic  $CC(P)$  and  $CC(P;Q)$  calculations for H<sub>2</sub>O carried out in this work, summarized in Table IV and Fig. 6. The authors of Ref. 118 obtained the CCSDTQ energies too, but we rely on our own CCSDTQ data, published in Ref. 63 and recalculated in this study, since Ref. 118 does not provide the CCSDTQ results for  $R_{O-H} = 2.5R_e$  and  $3R_e$  and the CCSDTQ energies for  $R_{O-H} = 1.5R_e$  and  $2R_e$  reported in Ref. 118 are in slight disagreement with the correctly converged values.

Up to twice the equilibrium O–H bond lengths, the CCSDT approach provides an accurate description of the electronic energies of water, resulting in the 0.493, 1.423, and  $-1.405$  millihartree signed errors relative to FCI at  $R_{O-H} = R_e, 1.5R_e,$  and  $2R_e$ , respectively, when the cc-pVDZ basis set is employed, but when  $R_{O-H} > 2R_e$ , CCSDT completely fails,<sup>63,118</sup> and the CCSD(T), CCSD(2)<sub>T</sub>, or CR-CC(2,3)<sub>A</sub> [in Table IV,  $\tau = 0$   $CC(P;Q)_{MP}$ ], CR-CC(2,3)<sub>D</sub> [in Table IV,  $\tau = 0$   $CC(P;Q)_{EN}$ ], CCSDt, and CC(t;3) approximations to CCSDT, which were examined in Refs. 63, 75, 88, and 118, fail with it [CCSD(T) fails already at  $R_{O-H} = 2R_e$ ]. In particular, the difference between the CCSDT and FCI energies obtained with the cc-pVDZ basis set at  $R_{O-H} = 2.5R_e$  is  $-24.752$  millihartree. At  $R_{O-H} = 3R_e$ , the situation becomes even more dramatic, with the CCSDT/cc-pVDZ energy falling 40.126 millihartree below its FCI counterpart.<sup>63,118</sup> One needs to incorporate  $T_4$  clusters to reduce these massive errors in the  $R_{O-H} > 2R_e$  region, and in order to do it in a reliable manner, one has to use full CCSDTQ or one of the robust approximations to it, such as the CCSDtq, CC(t,q;3), and CC(t,q;3,4) methods tested in Ref. 63. The conventional  $T_3$  plus  $T_4$  corrections to CCSD, such as CCSD(TQ)<sub>f</sub>,<sup>144</sup> or their CCSD(2)<sub>TQ</sub><sup>88,90</sup> and CR-CC(2,4)<sup>75,76,83</sup> counterparts examined in Refs. 63 and 88 do not suffice. The CCSDT(2)<sub>Q</sub> quadruples correction to CCSDT<sup>90</sup> is not robust enough either.<sup>88</sup>

When the cc-pVDZ basis set is employed, the differences between the CCSDTQ and FCI energies at  $R_{O-H} = R_e, 1.5R_e, 2R_e,$

$2.5R_e,$  and  $3R_e$  are 0.019, 0.121, 0.030,  $-2.361,$  and  $-4.733$  millihartree, respectively,<sup>63</sup> which is a huge improvement over CCSDT. One might argue the need for the inclusion of  $T_n$  clusters with  $n > 4$  at  $R_{O-H} = 2.5R_e$  and  $3R_e$  or try to obtain further improvements in describing the  $R_{O-H} > 2R_e$  region by replacing the RHF reference determinants used throughout this work by their unrestricted counterparts, but studies of this kind are outside the scope of this article. The goal of the calculations for the water molecule discussed in this subsection is to explore the potential offered by the semi-stochastic  $CC(P)$  and  $CC(P;Q)$  approaches, especially the  $CC(P;Q)$  corrections to the  $CC(P)$  energies calculated with the help of the FCIQMC and CISDTQ-MC propagations, in converging the CCSDTQ energetics obtained with the spin- and symmetry-adapted RHF references.

As demonstrated in Table IV and Fig. 6, which show the convergence of the  $CC(P)$  and  $CC(P;Q)$  energies toward their CCSDTQ parents, and Table S.4 of the supplementary material, which reports the total numbers of walkers characterizing the underlying CIQMC runs as percentages of the walker populations at  $\tau = 10.0$  a.u., where our CIQMC propagations were terminated, the semi-stochastic  $CC(P;Q)$  calculations using FCIQMC and CISDTQ-MC are extremely efficient in capturing the combined effects of  $T_3$  and  $T_4$  correlations. This remains true even in the most challenging  $R_{O-H} > 2R_e$  region, where the  $T_4$  contributions, which have to overcome the massive failures of the CCSDT approach, are very large and difficult to balance with their  $T_3$  counterparts. The FCIQMC- and CISDTQ-MC-driven  $CC(P;Q)$  computations accurately reproduce the parent CCSDTQ energetics already in the early stages of the underlying CIQMC propagations, when the stochastically determined  $P$  spaces contain small fractions of triples and even smaller fractions of quadruples and when the total numbers of walkers used in the CIQMC runs are much smaller than those required to converge these runs. The FCIQMC- and CISDTQ-MC-based  $CC(P;Q)$  approaches greatly accelerate convergence of the corresponding  $CC(P)$  calculations, despite that in our current implementation of the semi-stochastic  $CC(P;Q)$  routines aimed at CCSDTQ the noniterative correction  $\delta(P;Q)$  corrects the energy obtained by solving the  $CC(P)$  equations in the space of all singles and doubles and subsets of triples and quadruples captured by FCIQMC or CISDTQ-MC for the triples outside the stochastically determined  $P$  space, but not for the quadruples missed by CIQMC.

Similar to the previously discussed  $CC(P;Q)$  calculations aimed at CCSDT, the  $CC(P;Q)$  approach targeting CCSDTQ that adopts the  $CC(P;Q)_{EN}$  correction is generally most effective, although the results of the  $CC(P;Q)_{MP}$  calculations, in which the Epstein–Nesbet denominator  $D_K(P)$  in Eq. (9) is replaced by its Møller–Plesset form, are as accurate as their  $CC(P;Q)_{EN}$  counterparts in the quasi-degenerate  $R_{O-H} > 2R_e$  region. Indeed, when we look at the results in Table IV corresponding to  $R_{O-H} = 2.5R_e$  and  $3R_e$ , where the  $T_4$  effects, estimated by forming the differences of the CCSDTQ and CCSDT energies, exceed 22 and 35 millihartree, respectively,<sup>63</sup> and where the differences between the CCSDT and CCSD energies, which measure the magnitude of  $T_3$  contributions, are about  $-45$  and  $-51$  millihartree, respectively,<sup>63,118</sup> the FCIQMC- and CISDTQ-MC-based  $CC(P;Q)_{EN}$  computations reduce the large  $-20.739$  ( $R_{O-H} = 2.5R_e$ ) and  $-35.823$  ( $R_{O-H} = 3R_e$ ) millihartree errors relative to CCSDTQ obtained in the initial CR-CC(2,3)<sub>D</sub> [ $\tau = 0$   $CC(P;Q)_{EN}$ ] calculations to fractions of a millihartree after only 20 000  $\delta\tau = 0.0001$  a.u. MC iterations, i.e., after the FCIQMC

**TABLE IV.** Convergence of the  $CC(P)$ ,  $CC(P;Q)_{MP}$ , and  $CC(P;Q)_{EN}$  energies toward CCSDTQ, 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  $H_2O$  molecule, as described by the cc-pVDZ basis set, taken from Ref. 118. 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.

$R_{O-H}/R_e^a$	MC iterations	% of triples/quadruples		$CC(P)^b$		$CC(P;Q)_{MP}^b$		$CC(P;Q)_{EN}^b$	
		FCI <sup>c</sup>	CIQ <sup>d</sup>	FCI <sup>c</sup>	CIQ <sup>d</sup>	FCI <sup>c</sup>	CIQ <sup>d</sup>	FCI <sup>c</sup>	CIQ <sup>d</sup>
1.0	0		0/0	3.725 <sup>e</sup>		0.887 <sup>f</sup>		0.325 <sup>g</sup>	
	10 000	2/0	2/0	3.291	3.291	0.718	0.718	0.220	0.220
	20 000	4/1	4/1	2.874	2.874	0.633	0.629	0.205	0.185
	30 000	6/1	5/1	2.637	2.637	0.544	0.600	0.143	0.184
	40 000	11/2	9/2	2.052	2.052	0.441	0.471	0.142	0.129
	50 000	13/2	14/3	1.910	1.910	0.390	0.358	0.105	0.095
	60 000	17/3	18/4	1.481	1.481	0.304	0.323	0.087	0.106
	70 000	22/5	22/5	1.238	1.238	0.245	0.249	0.065	0.076
	80 000	27/6	27/6	0.956	0.956	0.207	0.216	0.073	0.082
	100 000	36/10	35/10	0.586	0.586	0.127	0.143	0.048	0.065
	$\infty$		100	-76.241 841 <sup>h</sup>		...		...	
1.5	0		0/0	9.922 <sup>e</sup>		2.704 <sup>f</sup>		1.021 <sup>g</sup>	
	10 000	3/1	3/1	6.612	6.545	1.393	1.501	0.290	0.434
	20 000	8/1	7/1	4.068	4.168	0.898	0.799	0.236	0.138
	30 000	11/2	11/2	3.000	3.032	0.613	0.698	0.144	0.248
	40 000	16/3	17/3	1.878	2.207	0.481	0.503	0.231	0.189
	50 000	22/4	22/4	1.465	1.507	0.377	0.366	0.185	0.166
	60 000	26/6	27/6	0.993	0.959	0.254	0.270	0.133	0.152
	70 000	31/8	33/9	0.786	0.706	0.229	0.206	0.133	0.122
	80 000	36/10	38/11	0.552	0.548	0.186	0.156	0.130	0.091
	100 000	46/17	48/18	0.259	0.263	0.086	0.086	0.061	0.060
	$\infty$		100	-76.072 227 <sup>h</sup>		...		...	
2.0	0		0/0	22.002 <sup>e</sup>		3.775 <sup>f</sup>		-0.581 <sup>g</sup>	
	10 000	2/0	2/0	11.766	11.803	1.966	2.189	-0.044	0.200
	20 000	7/1	6/1	4.172	4.937	1.129	1.295	0.567	0.626
	30 000	10/2	9/1	3.132	3.788	0.708	0.683	0.323	0.160
	40 000	14/3	13/2	1.728	1.966	0.603	0.668	0.436	0.483
	50 000	19/4	19/4	1.123	1.120	0.421	0.509	0.324	0.437
	60 000	25/6	24/6	0.794	0.719	0.305	0.221	0.246	0.156
	70 000	30/8	30/8	0.429	0.427	0.129	0.144	0.094	0.110
	80 000	36/11	35/11	0.327	0.293	0.106	0.103	0.079	0.082
	100 000	47/18	47/18	0.107	0.102	0.036	0.026	0.029	0.021
	$\infty$		100	-75.951 635 <sup>h</sup>		...		...	
2.5	0		0/0	22.668 <sup>e</sup>		-13.469 <sup>f</sup>		-20.739 <sup>g</sup>	
	10 000	3/0	3/0	18.305	-3.327	-1.136	-18.549	-4.962	-21.357
	20 000	6/1	6/1	5.254	7.207	0.010	0.448	-0.821	-0.588
	30 000	10/2	9/2	2.278	2.109	0.513	0.988	0.298	0.872
	40 000	15/3	13/3	1.021	1.170	0.304	0.542	0.220	0.490
	50 000	22/5	17/4	0.459	0.585	0.264	0.287	0.254	0.264
	60 000	27/8	23/6	0.340	0.424	0.105	0.222	0.096	0.212
	70 000	34/12	29/9	0.133	0.411	0.059	0.020	0.054	-0.033
	80 000	42/16	36/13	0.088	0.155	0.014	0.052	0.011	0.045
	100 000	55/28	49/22	0.020	0.027	0.013	0.020	0.012	0.020
	$\infty$		100	-75.920 352 <sup>h</sup>		...		...	

TABLE IV. (Continued.)

$R_{O-H}/R_e^a$	MC iterations	% of triples/quadruples		CC( $P$ ) <sup>b</sup>		CC( $P;Q$ ) <sub>MP</sub> <sup>b</sup>		CC( $P;Q$ ) <sub>EN</sub> <sup>b</sup>	
		FCI <sup>c</sup>	CIQ <sup>d</sup>	FCI <sup>c</sup>	CIQ <sup>d</sup>	FCI <sup>c</sup>	CIQ <sup>d</sup>	FCI <sup>c</sup>	CIQ <sup>d</sup>
3.0	0		0/0		15.582 <sup>e</sup>		-28.302 <sup>f</sup>		-35.823 <sup>g</sup>
	10 000	3/1	3/1	10.165	12.515	-2.390	-1.199	-3.945	-2.697
	20 000	5/1	5/1	4.282	2.721	-0.084	-0.690	-0.403	-0.875
	30 000	9/2	8/2	1.616	3.019	0.544	0.357	0.414	0.007
	40 000	13/3	11/3	0.969	0.830	0.267	0.378	0.199	0.334
	50 000	18/5	17/5	0.523	0.400	0.251	0.196	0.231	0.184
	60 000	24/8	22/7	0.185	0.237	0.097	0.093	0.090	0.087
	70 000	30/12	28/10	0.082	0.128	0.039	0.076	0.036	0.075
	80 000	36/16	34/14	0.030	0.050	0.022	0.030	0.021	0.029
	100 000	51/28	48/24	0.005	0.012	0.005	0.008	0.005	0.008
	∞		100		-75.916 679 <sup>h</sup>		...		...

<sup>a</sup>The equilibrium geometry,  $R_{O-H} = R_e$ , 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. 118.

<sup>b</sup>Unless otherwise stated, all energies are reported as errors relative to CCSDTQ in millihartree.

<sup>c</sup>FCI stands for *i*-FCIQMC.

<sup>d</sup>CIQ stands for *i*-CISDTQ-MC.

<sup>e</sup>Equivalent to CCSD.

<sup>f</sup>Equivalent to the CCSD energy corrected for the effects of  $T_3$  clusters using the CCSD(2)<sub>T</sub> approach of Ref. 88, which is equivalent to the approximate form of the completely renormalized CR-CC(2,3) approach of Refs. 75 and 76, abbreviated sometimes as CR-CC(2,3)<sub>A</sub> or CR-CC(2,3)<sub>A</sub>.<sup>62,63,78,80,83</sup>

<sup>g</sup>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. 75 and 76, abbreviated sometimes as CR-CC(2,3)<sub>D</sub> or CR-CC(2,3)<sub>D</sub>.<sup>62,63,78,80,83</sup>

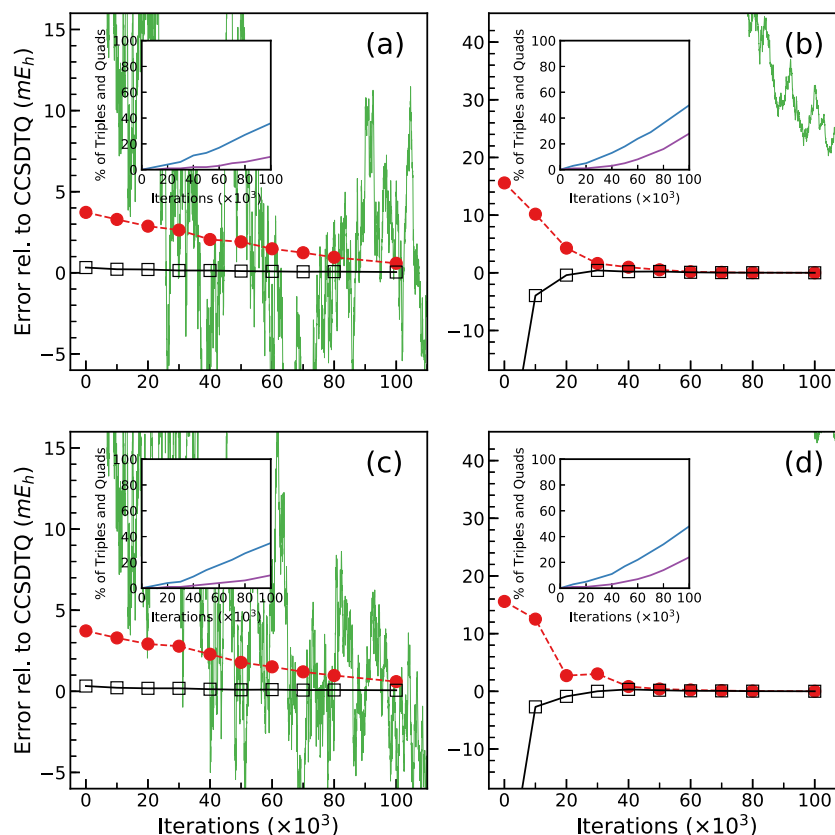
<sup>h</sup>Total CCSDTQ energy in hartree.

and CISDTQ-MC propagations capture as little as 5%–6% of the triples and 1% of the quadruples in the corresponding  $P$  spaces. The FCIQMC- and CISDTQ-MC-driven CC( $P;Q$ )<sub>MP</sub> calculations using the same QMC propagation time  $\tau$  are similarly effective though. They reduce the large -13.469 and -28.302 millihartree errors relative to CCSDTQ resulting from the initial CCSD(2)<sub>T</sub> or CR-CC(2,3)<sub>A</sub> [ $\tau = 0$  CC( $P;Q$ )<sub>MP</sub>] computations to a submillihartree level too.

The situation changes in the  $R_{O-H} = R_e - 2R_e$  region, where the  $T_4$  effects are much smaller than those originating from the  $T_3$  clusters. In this case, the convergence of the energies obtained in the semi-stochastic CC( $P;Q$ )<sub>MP</sub> calculations toward CCSDTQ is slower than that obtained with the CC( $P;Q$ )<sub>EN</sub> approach, i.e., our earlier conclusion, drawn from the calculations discussed in Secs. III A and III B and Ref. 58, that the use of the CC( $P;Q$ )<sub>EN</sub> corrections to the semi-stochastic CC( $P$ ) energies is generally most effective still stands. This becomes particularly clear when we compare the results of the FCIQMC- and CISDTQ-MC-driven CC( $P;Q$ )<sub>MP</sub> and CC( $P;Q$ )<sub>EN</sub> calculations at  $R_{O-H} = R_e$  and  $1.5R_e$ . For example, it takes only 40 000  $\delta\tau = 0.0001$  a.u. MC time steps, or about 10% of the triples and 2% of the quadruples captured in the  $P$  space, for the CC( $P;Q$ )<sub>EN</sub> approach to reach a 0.1 millihartree accuracy level relative to CCSDTQ at  $R_{O-H} = R_e$ . The CC( $P;Q$ )<sub>MP</sub> calculations reach the same accuracy level after 100 000 MC time steps that capture about 35% of the triples and 10% of the quadruples. When the  $R_{O-H} = 1.5R_e$  geometry is considered, the CC( $P;Q$ )<sub>EN</sub> calculations reach a 0.1 millihartree accuracy level relative to CCSDTQ after 60 000–70 000 MC iterations that capture about 30% of the triples

and 6%–9% of the quadruples, i.e., in the relatively early stages of the FCIQMC and CISDTQ-MC propagations. The CC( $P;Q$ )<sub>MP</sub> calculations reach a similar accuracy level 20 000–30 000 MC iterations later, after capturing about 40% of the triples and more than 10% of the quadruples. It is certainly reassuring that the CC( $P;Q$ )<sub>EN</sub> calculations using FCIQMC and CISDTQ-MC to identify the leading triply and quadruply excited determinants for the inclusion in the underlying  $P$  spaces are capable of reproducing the CCSDTQ energies of the water molecule over a wide range of geometries along the  $C_{2v}$ -symmetric cut of the ground-state potential energy surface considered in Table IV and Fig. 6 to within ~0.1 millihartree out of the early stages of the CIQMC propagations, after capturing about 10% ( $R_{O-H} = R_e$ ) or 30% ( $R_{O-H} > R_e$ ) of the triples and 2% ( $R_{O-H} = R_e$ ) or about 10% ( $R_{O-H} > R_e$ ) of the quadruples. Having said this, it is interesting to observe that both types of the CC( $P;Q$ ) corrections tested in this study, abbreviated as CC( $P;Q$ )<sub>MP</sub> and CC( $P;Q$ )<sub>EN</sub>, perform equally well when  $R_{O-H} > 2R_e$ , i.e., when the  $T_3$  and  $T_4$  effects are both very large. We observed a similar behavior in Ref. 63, when examining the relative performance of the CC( $P;Q$ )-based CC( $t,q;3$ )<sub>A</sub> and CC( $t,q;3$ )<sub>D</sub> corrections to CCSDTQ using the double dissociation of water as one of the examples. This should not be surprising since the CC( $t,q;3$ )<sub>A</sub> and CC( $t,q;3$ )<sub>D</sub> methods investigated in Ref. 63 can be regarded as the deterministic counterparts of the semi-stochastic CC( $P;Q$ )<sub>MP</sub> and CC( $P;Q$ )<sub>EN</sub> approaches targeting the CCSDTQ energetics implemented in this work.

As in the case of the CC( $P;Q$ ) calculations targeting CCSDT, discussed in Secs. III A and III B, the observed fast convergence of the semi-stochastic CC( $P;Q$ ) calculations aimed at recovering the



**FIG. 6.** Convergence of the  $CC(P)$  (red filled circles and dashed lines) and  $CC(P;Q)_{EN}$  (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 captured by  $i$ -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 captured 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  $\angle(\text{H–O–H})$  angle. All reported energies are errors relative to CCSDTQ in millihartree. The insets show the percentages of triples (blue line) and quadruples (purple line) captured during the relevant  $i$ -CIQMC propagations.

CCSDTQ energetics does not seem to be affected by the type of the CIQMC approach used to identify the leading triply and quadruply excited determinants. This should facilitate future applications of the semi-stochastic  $CC(P;Q)$  methodology, including cases of stronger electronic quasi-degeneracies characterized by large  $T_3$  and  $T_4$  contributions, helping us to converge the CCSDTQ-level energetics at the small fraction of the deterministic CCSDTQ effort by taking advantage of the least expensive forms CIQMC capable of capturing triples and quadruples, represented in this study by CISDTQ-MC.

#### IV. CONCLUSIONS

We have recently started exploring a novel way of obtaining accurate electronic energetics equivalent to high-level CC calculations at the small fraction of the computational effort and preserving the black-box character of conventional single-reference computations by merging the deterministic  $CC(P;Q)$  formalism,

originally proposed in Refs. 57 and 61, along with the underlying  $CC(P)/EOMCC(P)$  framework, with the stochastic CIQMC<sup>64–67</sup> and  $CCMC$ <sup>68–71</sup> approaches.<sup>58–60</sup> When combined with the FCIQMC and CCSDT-MC wave function sampling, used to identify the leading triply excited determinants or cluster/excitation amplitudes, and correcting the  $CC(P)$ <sup>58</sup> and  $EOMCC(P)$ <sup>59</sup> energies for the remaining triples not captured by FCIQMC or CCSDT-MC, the resulting semi-stochastic  $CC(P;Q)$  methodology<sup>58</sup> and its excited-state extension<sup>60</sup> turned out to be very promising, allowing us to converge the CCSDT and  $EOMCCSDT$  energetics out of the early stages of the underlying QMC propagations.

This study can be regarded as the next key step in the development and exploration of the semi-stochastic  $CC(P;Q)$  approaches, in which we have extended our initial work,<sup>58</sup> focusing on recovering the CCSDT energetics and relying on FCIQMC and CCSDT-MC, to more efficient ways of identifying the leading higher-than-doubly excited determinants for the inclusion in the underlying

$P$  spaces. We have accomplished this goal by replacing FCIQMC by its less expensive CISDT-MC and CISDTQ-MC counterparts. We have also developed and tested the initial variant of the semi-stochastic  $CC(P;Q)$  method aimed at converging the CCSDTQ energetics, in which the results of  $CC(P)$  calculations in the subspaces spanned by singles, doubles, and subsets of triples and quadruples identified by FCIQMC or CISDTQ-MC are corrected for the remaining triples outside the stochastically determined  $P$  spaces. By comparing the FCIQMC-driven  $CC(P;Q)$  calculations targeting CCSDT, carried out in this work, in which the noniterative corrections  $\delta(P;Q)$  to the  $CC(P)$  energies have been treated fully, as required by Eqs. (8), (9), and (11), with the analogous computations reported in Ref. 58, where the same corrections were treated in a somewhat simplified manner by neglecting the three-body component of the de-excitation operator  $\Lambda^{(P)}$  used to construct amplitudes  $\ell_K(P)$  entering Eq. (8), we have examined the significance of the full vs approximate treatment of these corrections for the accuracy of the resulting  $CC(P;Q)$  energies. Other important issues, such as the benefits of using the Epstein–Nesbet form of the denominators  $D_K(P)$  that enter the definition of corrections  $\delta(P;Q)$ , resulting in the  $CC(P;Q)_{\text{EN}}$  variant of  $CC(P;Q)$ , as compared to their Møller–Plesset counterparts defining the  $CC(P;Q)_{\text{MP}}$  corrections, have been investigated as well.

The ability of the semi-stochastic  $CC(P;Q)$  approaches to converge the CCSDT and CCSDTQ energies based on the truncated CISDT-MC and CISDTQ-MC propagations and their FCIQMC counterparts, in which the noniterative corrections  $\delta(P;Q)$  have been treated fully, has been illustrated using a few molecular examples, for which the deterministic CCSDT and CCSDTQ calculations that provide the reference data are feasible and which require a high-level CC treatment to obtain a reliable description. Thus, we have reported the results of the semi-stochastic  $CC(P;Q)$  calculations using CISDT-MC, CISDTQ-MC, and FCIQMC aimed at converging the CCSDT energetics for the F–F bond breaking in  $F_2$  and the automerization of cyclobutadiene, which require an accurate treatment of  $T_3$  clusters accounting for the relaxation of  $T_1$  and  $T_2$  amplitudes in the presence of large  $T_3$  contributions, and the CISDTQ-MC- and FCIQMC-driven  $CC(P;Q)$  computations for the  $C_{2v}$ -symmetric stretching of the O–H bonds in the water molecule targeting CCSDTQ, where the  $T_3$  and  $T_4$  clusters become large and difficult to balance.

The numerical results reported in this article clearly show that the semi-stochastic  $CC(P;Q)$  calculations are capable of accurately reproducing the parent CCSDT and CCSDTQ energetics, even when electronic quasi-degeneracies and higher-than-two-body components of the cluster operator become large, out of the early stages of the corresponding CIQMC propagations, accelerating convergence of the underlying  $CC(P)$  computations at the same time. The convergence of the  $CC(P;Q)$  energies toward their CCSDT and CCSDTQ parents does not seem to be affected by the type of the CIQMC approach used to identify the leading triply or triply and quadruply excited determinants. In the case of the  $CC(P;Q)$  calculations targeting the CCSDT energetics, one can use FCIQMC or one of its less expensive truncated forms, such as CISDTQ-MC, or even the crude CISDT-MC approach, with virtually no impact on the systematic convergence pattern toward CCSDT as the propagation time  $\tau$  approaches  $\infty$ . Similarly, one can replace FCIQMC by CISDTQ-MC without any significant effect on the convergence

of the semi-stochastic  $CC(P;Q)$  calculations toward CCSDTQ. Our calculations also suggest that a complete treatment of the  $CC(P;Q)$  corrections  $\delta(P;Q)$ , as defined by Eqs. (8), (9), and (11), including the use of the  $CC(P;Q)_{\text{EN}}$  approach, as opposed to its more approximate  $CC(P;Q)_{\text{MP}}$  version, is more important than the actual type of the CIQMC approach used to determine the relevant  $P$  spaces, especially when one is interested in accelerating convergence of the semi-stochastic  $CC(P;Q)$  calculations in the early stages of the QMC propagations. We have demonstrated that independent of the type of the CIQMC approach used to identify the leading triply or triply and quadruply excited determinants for the inclusion in the relevant  $P$  spaces and independent of the magnitude of  $T_3$  and  $T_4$  effects, the semi-stochastic  $CC(P;Q)$  calculations allow us to reach submillihartree accuracy levels relative to the parent CCSDT and CCSDTQ energetics with small fractions of higher-than-doubly excited determinants captured in the early stages of the corresponding CIQMC runs and with small walker populations that are far less than the total numbers of walkers required to converge these runs.

By relaxing  $T_1$  and  $T_2$  clusters in the presence of their  $T_3$  or  $T_3$  and  $T_4$  counterparts defined using the excitation lists provided by full or truncated CIQMC, the semi-stochastic  $CC(P;Q)$  computations are capable of considerably improving accuracy of the more established noniterative corrections to CCSD without making the calculations a lot more expensive. In this sense, the semi-stochastic  $CC(P;Q)$  methodology using CIQMC is very similar to the deterministic  $CC(t;3)$ ,  $CC(t,q;3)$ , and  $CC(t,q;3,4)$  hierarchy developed and tested in Refs. 57, 61–63, 82, and 100, which uses the  $CC(P;Q)$  corrections to correct the results of the active-space CCSDt or CCSDtq calculations for the remaining  $T_3$  or  $T_3$  and  $T_4$  correlations that were not captured via active orbitals. There is, however, one major advantage of the semi-stochastic  $CC(P;Q)$  framework over the  $CC(t;3)$ ,  $CC(t,q;3)$ , and  $CC(t,q;3,4)$  approaches, namely, the use of FCIQMC or truncated CIQMC propagations, which can efficiently identify the leading higher-than-doubly excited determinants for the inclusion in the relevant  $P$  spaces, combined with the  $\delta(P;Q)$  corrections to capture the remaining correlations of interest, offers an automated way of performing accurate  $CC(P;Q)$  computations without any reference to the user- and system-dependent active orbitals. The analogies between the active-space CCSDt (for excited states, EOMCCSDt<sup>26,27,145</sup>) and semi-stochastic  $CC(P)/\text{EOMCC}(P)$  approaches, on which the deterministic  $CC(t;3)$  (in the case of CCSDt/EOMCCSDt) and CIQMC-driven [in the case of semi-stochastic  $CC(P)/\text{EOMCC}(P)$ ]  $CC(P;Q)$  approaches are based, have been investigated in Ref. 60.

The findings presented in this article are encouraging from the point of view of future applications of the semi-stochastic  $CC(P;Q)$  methodology using CIQMC, including challenging cases of stronger electronic quasi-degeneracies characterized by large  $T_3$  or  $T_3$  and  $T_4$  contributions that other approximations to CCSDT or CCSDTQ may struggle with, but the story is not over yet. We certainly need to improve the efficiency of our  $CC(P;Q)$  codes, especially the underlying  $CC(P)$  routines, to obtain full benefits offered by the semi-stochastic  $CC(P;Q)$  approaches, discussed in Sec. II B. This is especially true in the case of our current  $CC(P;Q)$  codes aimed at converging the CCSDTQ energetics, which have a largely pilot character. In this case, we also need to examine if one can further improve the convergence of the FCIQMC- or CISDTQ-MC-driven  $CC(P;Q)$  calculations aimed at CCSDTQ by correcting the underlying  $CC(P)$

energies for both the missing triples and quadruples not captured by CIQMC at a given time  $\tau$ , not just for the missing triples, as has been done in this work. It would also be useful to examine if one can extend the semi-stochastic  $CC(P)$  and  $CC(P;Q)$  approaches to the higher CC theory levels, beyond CCSDTQ examined in this work and beyond EOMCCSDT explored in Refs. 59 and 60, and investigate if our observations regarding the utility of the truncated CIQMC methods, such as CISDT-MC and CISDTQ-MC, remain true in the excited-state and open-shell  $CC(P;Q)$  calculations. In this study, we have adopted the original form of the  $i$ -CIQMC algorithm proposed in Ref. 65, but it would be interesting to examine if one could obtain additional benefits by interfacing our semi-stochastic  $CC(P;Q)$  methods with the improved ways of converging CIQMC, such as the adaptive-shift approach developed in Refs. 67 and 114. All of the above ideas are presently pursued in our group, and the results will be reported as soon as they become available. Last but not least, we have recently interfaced our  $CC(P)$  and  $CC(P;Q)$  routines with some of the modern versions of the selected CI approaches, which date back to the late 1960s and early 1970s<sup>146–149</sup> and which have recently regained significant attention.<sup>150–160</sup> Our initial numerical results, which we hope to report in a separate publication,<sup>161</sup> indicate that selected CI methods can be as effective in generating meaningful  $P$  spaces for the  $CC(P)$  calculations, which precede the determination of the  $\delta(P;Q)$  moment corrections, as the stochastic CIQMC propagations advocated in this and our earlier<sup>58–60</sup> studies.

## SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for the information about the total numbers of walkers characterizing the FCIQMC, CISDT-MC, and CISDTQ-MC propagations for the F–F bond breaking in  $F_2$  and the automerization of cyclobutadiene and the FCIQMC and CISDTQ-MC propagations for the double dissociation of the water molecule carried out in the present study.

## ACKNOWLEDGMENTS

This work was supported by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy (Grant No. DE-FG02-01ER15228 to P.P.); the National Science Foundation (Grant No. CHE-1763371 to P.P.); and Phase I and II Software Fellowships awarded to J.E.D. by the Molecular Sciences Software Institute funded by the National Science Foundation (Grant No. ACI-1547580). P.P. thanks Professor Ali Alavi, Professor George H. Booth, and Professor Alex J. W. Thom for useful discussions.

## DATA AVAILABILITY

The data that support the findings of this study are available within the article and its [supplementary material](#).

## REFERENCES

- J. Hubbard, *Proc. R. Soc. London, Ser. A* **240**, 539 (1957).
- N. M. Hugenholtz, *Physica* **23**, 481 (1957).
- F. Coester, *Nucl. Phys.* **7**, 421 (1958).
- F. Coester and H. Kümmel, *Nucl. Phys.* **17**, 477 (1960).
- J. Čížek, *J. Chem. Phys.* **45**, 4256 (1966).
- J. Čížek, *Adv. Chem. Phys.* **14**, 35 (1969).
- J. Paldus, J. Čížek, and I. Shavitt, *Phys. Rev. A* **5**, 50 (1972).
- J. Paldus and X. Li, *Adv. Chem. Phys.* **110**, 1 (1999).
- P. Piecuch and K. Kowalski, *Int. J. Mol. Sci.* **3**, 676 (2002).
- R. J. Bartlett and M. Musiał, *Rev. Mod. Phys.* **79**, 291 (2007).
- D. I. Lyakh, M. Musiał, V. F. Lotrich, and R. J. Bartlett, *Chem. Rev.* **112**, 182 (2012).
- F. A. Evangelista, *J. Chem. Phys.* **149**, 030901 (2018).
- G. D. Purvis III and R. J. Bartlett, *J. Chem. Phys.* **76**, 1910 (1982).
- J. M. Cullen and M. C. Zerner, *J. Chem. Phys.* **77**, 4088 (1982).
- G. E. Scuseria, A. C. Scheiner, T. J. Lee, J. E. Rice, and H. F. Schaefer III, *J. Chem. Phys.* **86**, 2881 (1987).
- P. Piecuch and J. Paldus, *Int. J. Quantum Chem.* **36**, 429 (1989).
- J. Noga and R. J. Bartlett, *J. Chem. Phys.* **86**, 7041 (1987); Erratum **89**, 3401 (1988).
- G. E. Scuseria and H. F. Schaefer III, *Chem. Phys. Lett.* **152**, 382 (1988).
- J. D. Watts and R. J. Bartlett, *J. Chem. Phys.* **93**, 6104 (1990).
- N. Oliphant and L. Adamowicz, *J. Chem. Phys.* **95**, 6645 (1991).
- S. A. Kucharski and R. J. Bartlett, *Theor. Chim. Acta* **80**, 387 (1991).
- S. A. Kucharski and R. J. Bartlett, *J. Chem. Phys.* **97**, 4282 (1992).
- K. Emrich, *Nucl. Phys. A* **351**, 379 (1981).
- J. Geertsen, M. Rittby, and R. J. Bartlett, *Chem. Phys. Lett.* **164**, 57 (1989).
- J. F. Stanton and R. J. Bartlett, *J. Chem. Phys.* **98**, 7029 (1993).
- K. Kowalski and P. Piecuch, *J. Chem. Phys.* **115**, 643 (2001).
- K. Kowalski and P. Piecuch, *Chem. Phys. Lett.* **347**, 237 (2001).
- S. A. Kucharski, M. Włoch, M. Musiał, and R. J. Bartlett, *J. Chem. Phys.* **115**, 8263 (2001).
- M. Kállay and J. Gauss, *J. Chem. Phys.* **121**, 9257 (2004).
- S. Hirata, *J. Chem. Phys.* **121**, 51 (2004).
- H. J. Monkhorst, *Int. J. Quantum Chem.* **12**, 421 (1977).
- E. Dalgaard and H. J. Monkhorst, *Phys. Rev. A* **28**, 1217 (1983).
- D. Mukherjee and P. K. Mukherjee, *Chem. Phys.* **39**, 325 (1979).
- H. Sekino and R. J. Bartlett, *Int. J. Quantum Chem.* **26**, 255 (1984).
- M. Takahashi and J. Paldus, *J. Chem. Phys.* **85**, 1486 (1986).
- H. Koch and P. Jørgensen, *J. Chem. Phys.* **93**, 3333 (1990).
- H. Koch, H. J. A. Jensen, P. Jørgensen, and T. Helgaker, *J. Chem. Phys.* **93**, 3345 (1990).
- A. E. Kondo, P. Piecuch, and J. Paldus, *J. Chem. Phys.* **102**, 6511 (1995).
- A. E. Kondo, P. Piecuch, and J. Paldus, *J. Chem. Phys.* **104**, 8566 (1995).
- M. Urban, J. Noga, S. J. Cole, and R. J. Bartlett, *J. Chem. Phys.* **83**, 4041 (1985).
- K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, *Chem. Phys. Lett.* **157**, 479 (1989).
- Y. S. Lee and R. J. Bartlett, *J. Chem. Phys.* **80**, 4371 (1984).
- Y. S. Lee, S. A. Kucharski, and R. J. Bartlett, *J. Chem. Phys.* **81**, 5906 (1984); Erratum **82**, 5761 (1985).
- H. Koch, O. Christiansen, P. Jørgensen, and J. Olsen, *Chem. Phys. Lett.* **244**, 75 (1995).
- O. Christiansen, H. Koch, and P. Jørgensen, *J. Chem. Phys.* **103**, 7429 (1995).
- P. Piecuch, K. Kowalski, I. S. O. Pimienta, and M. J. McGuire, *Int. Rev. Phys. Chem.* **21**, 527 (2002).
- 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).
- P. Piecuch, *Mol. Phys.* **108**, 2987 (2010).
- K. Kowalski, D. J. Dean, M. Hjorth-Jensen, T. Papenbrock, and P. Piecuch, *Phys. Rev. Lett.* **92**, 132501 (2004).
- M. Włoch, D. J. Dean, J. R. Gour, M. Hjorth-Jensen, K. Kowalski, T. Papenbrock, and P. Piecuch, *Phys. Rev. Lett.* **94**, 212501 (2005).
- J. R. Gour, P. Piecuch, M. Hjorth-Jensen, M. Włoch, and D. J. Dean, *Phys. Rev. C* **74**, 024310 (2006).



- <sup>52</sup>M. Horoi, J. R. Gour, M. Włoch, M. D. Lodriguito, B. A. Brown, and P. Piecuch, *Phys. Rev. Lett.* **98**, 112501 (2007).
- <sup>53</sup>J. R. Gour, M. Horoi, P. Piecuch, and B. A. Brown, *Phys. Rev. Lett.* **101**, 052501 (2008).
- <sup>54</sup>R. Roth, J. R. Gour, and P. Piecuch, *Phys. Rev. C* **79**, 054325 (2009).
- <sup>55</sup>S. Binder, P. Piecuch, A. Calci, J. Langhammer, P. Navrátil, and R. Roth, *Phys. Rev. C* **88**, 054319 (2013).
- <sup>56</sup>G. Hagen, T. Papenbrock, M. Hjorth-Jensen, and D. J. Dean, *Rep. Prog. Phys.* **77**, 096302 (2014).
- <sup>57</sup>J. Shen and P. Piecuch, *Chem. Phys.* **401**, 180 (2012).
- <sup>58</sup>J. E. Deustua, J. Shen, and P. Piecuch, *Phys. Rev. Lett.* **119**, 223003 (2017).
- <sup>59</sup>J. E. Deustua, S. H. Yuwono, J. Shen, and P. Piecuch, *J. Chem. Phys.* **150**, 111101 (2019).
- <sup>60</sup>S. H. Yuwono, A. Chakraborty, J. Emiliano Deustua, J. Shen, and P. Piecuch, *Mol. Phys.* **118**, e1817592 (2020).
- <sup>61</sup>J. Shen and P. Piecuch, *J. Chem. Phys.* **136**, 144104 (2012).
- <sup>62</sup>J. Shen and P. Piecuch, *J. Chem. Theory Comput.* **8**, 4968 (2012).
- <sup>63</sup>N. P. Bauman, J. Shen, and P. Piecuch, *Mol. Phys.* **115**, 2860 (2017).
- <sup>64</sup>G. H. Booth, A. J. W. Thom, and A. Alavi, *J. Chem. Phys.* **131**, 054106 (2009).
- <sup>65</sup>D. Cleland, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **132**, 041103 (2010).
- <sup>66</sup>W. Dobrautz, S. D. Smart, and A. Alavi, *J. Chem. Phys.* **151**, 094104 (2019).
- <sup>67</sup>K. Ghanem, A. Y. Lozovoi, and A. Alavi, *J. Chem. Phys.* **151**, 224108 (2019).
- <sup>68</sup>A. J. W. Thom, *Phys. Rev. Lett.* **105**, 263004 (2010).
- <sup>69</sup>R. S. T. Franklin, J. S. Spencer, A. Zocante, and A. J. W. Thom, *J. Chem. Phys.* **144**, 044111 (2016).
- <sup>70</sup>J. S. Spencer and A. J. W. Thom, *J. Chem. Phys.* **144**, 084108 (2016).
- <sup>71</sup>C. J. C. Scott and A. J. W. Thom, *J. Chem. Phys.* **147**, 124105 (2017).
- <sup>72</sup>J. E. Deustua, I. Magoulas, J. Shen, and P. Piecuch, *J. Chem. Phys.* **149**, 151101 (2018).
- <sup>73</sup>E. Vitale, A. Alavi, and D. Kats, *J. Chem. Theory Comput.* **16**, 5621 (2020).
- <sup>74</sup>J. E. Eriksen, T. A. Anderson, J. E. Deustua, K. Ghanem, D. Hait, M. R. Hoffmann, S. Lee, D. S. Levine, I. Magoulas, J. Shen, N. M. Tubman, K. B. Whaley, E. Xu, Y. Yao, N. Zhang, A. Alavi, G. K.-L. Chan, M. Head-Gordon, W. Liu, P. Piecuch, S. Sharma, S. L. Ten-no, C. J. Umrigar, and J. Gauss, *J. Phys. Chem. Lett.* **11**, 8922 (2020).
- <sup>75</sup>P. Piecuch and M. Włoch, *J. Chem. Phys.* **123**, 224105 (2005).
- <sup>76</sup>P. Piecuch, M. Włoch, J. R. Gour, and A. Kinal, *Chem. Phys. Lett.* **418**, 467 (2006).
- <sup>77</sup>M. Włoch, M. D. Lodriguito, P. Piecuch, and J. R. Gour, *Mol. Phys.* **104**, 2149 (2006); Erratum **104**, 2991 (2006).
- <sup>78</sup>M. Włoch, J. R. Gour, and P. Piecuch, *J. Phys. Chem. A* **111**, 11359 (2007).
- <sup>79</sup>P. Piecuch, J. R. Gour, and M. Włoch, *Int. J. Quantum Chem.* **108**, 2128 (2008).
- <sup>80</sup>P. Piecuch, J. R. Gour, and M. Włoch, *Int. J. Quantum Chem.* **109**, 3268 (2009).
- <sup>81</sup>G. Fradelos, J. J. Lutz, T. A. Wesolowski, P. Piecuch, and M. Włoch, *J. Chem. Theory Comput.* **7**, 1647 (2011).
- <sup>82</sup>I. Magoulas, N. P. Bauman, J. Shen, and P. Piecuch, *J. Phys. Chem. A* **122**, 1350 (2018).
- <sup>83</sup>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.
- <sup>84</sup>P. Piecuch, M. Włoch, and A. J. C. Varandas, *Theor. Chem. Acc.* **120**, 59 (2008).
- <sup>85</sup>K. Jankowski, J. Paldus, and P. Piecuch, *Theor. Chim. Acta* **80**, 223 (1991).
- <sup>86</sup>P. Piecuch and K. Kowalski, in *Computational Chemistry: Reviews of Current Trends*, edited by J. Leszczyński (World Scientific, Singapore, 2000), Vol. 5, pp. 1–104.
- <sup>87</sup>K. Kowalski and P. Piecuch, *J. Chem. Phys.* **113**, 18 (2000).
- <sup>88</sup>S. Hirata, P.-D. Fan, A. A. Auer, M. Nooijen, and P. Piecuch, *J. Chem. Phys.* **121**, 12197 (2004).
- <sup>89</sup>S. R. Gwaltney and M. Head-Gordon, *Chem. Phys. Lett.* **323**, 21 (2000).
- <sup>90</sup>S. Hirata, M. Nooijen, I. Grabowski, and R. J. Bartlett, *J. Chem. Phys.* **114**, 3919 (2001); Erratum **115**, 3967 (2001).
- <sup>91</sup>S. R. Gwaltney and M. Head-Gordon, *J. Chem. Phys.* **115**, 2014 (2001).
- <sup>92</sup>J. F. Stanton, *Chem. Phys. Lett.* **281**, 130 (1997).
- <sup>93</sup>T. D. Crawford and J. F. Stanton, *Int. J. Quantum Chem.* **70**, 601 (1998).
- <sup>94</sup>S. A. Kucharski and R. J. Bartlett, *J. Chem. Phys.* **108**, 5243 (1998).
- <sup>95</sup>A. G. Taube and R. J. Bartlett, *J. Chem. Phys.* **128**, 044110 (2008).
- <sup>96</sup>A. G. Taube and R. J. Bartlett, *J. Chem. Phys.* **128**, 044111 (2008).
- <sup>97</sup>K. Kowalski and P. Piecuch, *J. Chem. Phys.* **122**, 074107 (2005).
- <sup>98</sup>Y. Ge, M. S. Gordon, and P. Piecuch, *J. Chem. Phys.* **127**, 174106 (2007).
- <sup>99</sup>Y. Ge, M. S. Gordon, P. Piecuch, M. Włoch, and J. R. Gour, *J. Phys. Chem. A* **112**, 11873 (2008).
- <sup>100</sup>S. H. Yuwono, I. Magoulas, J. Shen, and P. Piecuch, *Mol. Phys.* **117**, 1486 (2019).
- <sup>101</sup>N. Oliphant and L. Adamowicz, *J. Chem. Phys.* **96**, 3739 (1992).
- <sup>102</sup>P. Piecuch, N. Oliphant, and L. Adamowicz, *J. Chem. Phys.* **103**, 1875 (1993).
- <sup>103</sup>P. Piecuch and L. Adamowicz, *J. Chem. Phys.* **100**, 5792 (1994).
- <sup>104</sup>P. Piecuch and L. Adamowicz, *J. Chem. Phys.* **102**, 898 (1995).
- <sup>105</sup>K. B. Ghose, P. Piecuch, and L. Adamowicz, *J. Chem. Phys.* **103**, 9331 (1995).
- <sup>106</sup>L. Adamowicz, P. Piecuch, and K. B. Ghose, *Mol. Phys.* **94**, 225 (1998).
- <sup>107</sup>P. Piecuch, S. A. Kucharski, and R. J. Bartlett, *J. Chem. Phys.* **110**, 6103 (1999).
- <sup>108</sup>P. Piecuch, S. A. Kucharski, and V. Špirko, *J. Chem. Phys.* **111**, 6679 (1999).
- <sup>109</sup>N. Oliphant and L. Adamowicz, *J. Chem. Phys.* **94**, 1229 (1991).
- <sup>110</sup>M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, *J. Comput. Chem.* **14**, 1347 (1993).
- <sup>111</sup>G. M. J. Barca, C. Bertoni, L. Carrington, D. Datta, N. De Silva, J. E. Deustua, D. G. Fedorov, J. R. Gour, A. O. Gunina, E. Guidez, T. Harville, S. Irle, J. Ivanic, K. Kowalski, S. S. Leang, H. Li, W. Li, J. J. Lutz, I. Magoulas, J. Mato, V. Mironov, H. Nakata, B. Q. Pham, P. Piecuch, D. Poole, S. R. Pruitt, A. P. Rendell, L. B. Roskop, K. Ruedenberg, T. Sattasathuchana, M. W. Schmidt, J. Shen, L. Slipchenko, M. Sosonkina, V. Sundriyal, A. Tiwari, J. L. Galvez Vallejo, B. Westheimer, M. Włoch, P. Xu, F. Zahariev, and M. S. Gordon, *J. Chem. Phys.* **152**, 154102 (2020).
- <sup>112</sup>J. S. Spencer, N. S. Blunt, W. A. Vigor, F. D. Malone, W. M. C. Foulkes, J. J. Shepherd, and A. J. W. Thom, *J. Open Res. Software* **3**, e9 (2015).
- <sup>113</sup>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).
- <sup>114</sup>K. Ghanem, K. Guthrie, and A. Alavi, *J. Chem. Phys.* **153**, 224115 (2020).
- <sup>115</sup>K. Kowalski and P. Piecuch, *Chem. Phys. Lett.* **344**, 165 (2001).
- <sup>116</sup>A. Balková and R. J. Bartlett, *J. Chem. Phys.* **101**, 8972 (1994).
- <sup>117</sup>D. I. Lyakh, V. F. Lotrich, and R. J. Bartlett, *Chem. Phys. Lett.* **501**, 166 (2011).
- <sup>118</sup>J. Olsen, P. Jørgensen, H. Koch, A. Balkova, and R. J. Bartlett, *J. Chem. Phys.* **104**, 8007 (1996).
- <sup>119</sup>T. H. Dunning, Jr., *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>120</sup>R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).
- <sup>121</sup>J. Shen, E. Xu, Z. Kou, and S. Li, *J. Chem. Phys.* **132**, 114115 (2010).
- <sup>122</sup>J. Shen, Z. Kou, E. Xu, and S. Li, *J. Chem. Phys.* **133**, 234106 (2010).
- <sup>123</sup>J. Shen, Z. Kou, E. Xu, and S. Li, *J. Chem. Phys.* **134**, 044134 (2011).
- <sup>124</sup>D. W. Whitman and B. K. Carpenter, *J. Am. Chem. Soc.* **104**, 6473 (1982).
- <sup>125</sup>B. A. Hess, P. Čársky, and L. J. Schaad, *J. Am. Chem. Soc.* **105**, 695 (1983).
- <sup>126</sup>B. K. Carpenter, *J. Am. Chem. Soc.* **105**, 1700 (1983).
- <sup>127</sup>A. F. Voter and W. A. Goddard III, *J. Am. Chem. Soc.* **108**, 2830 (1986).
- <sup>128</sup>P. Čársky, R. J. Bartlett, G. Fitzgerald, J. Nova, and V. Špirko, *J. Chem. Phys.* **89**, 3008 (1988).
- <sup>129</sup>B. R. Arnold and J. Michl, *J. Phys. Chem.* **97**, 13348 (1993).
- <sup>130</sup>W. Wu, Y. Mo, Z. Cao, and Q. Zhang, *Theor. Comput. Chem.* **10**, 143 (2002).
- <sup>131</sup>S. V. Levchenko and A. I. Krylov, *J. Chem. Phys.* **120**, 175 (2004).
- <sup>132</sup>O. Demel and J. Pittner, *J. Chem. Phys.* **124**, 144112 (2006).
- <sup>133</sup>S. Saddique and G. A. Worth, *Chem. Phys.* **329**, 99 (2006).
- <sup>134</sup>M. Eckert-Maksić, M. Vazdar, M. Barbatti, H. Lischka, and Z. B. Maksić, *J. Chem. Phys.* **125**, 064310 (2006).

- <sup>135</sup>K. Bhaskaran-Nair, O. Demel, and J. Pittner, *J. Chem. Phys.* **129**, 184105 (2008).
- <sup>136</sup>P. B. Karadakov, *J. Phys. Chem. A* **112**, 7303 (2008).
- <sup>137</sup>O. Demel, K. R. Shamasundar, L. Kong, and M. Nooijen, *J. Phys. Chem. A* **112**, 11895 (2008).
- <sup>138</sup>J. Shen, T. Fang, S. Li, and Y. Jiang, *J. Phys. Chem. A* **112**, 12518 (2008).
- <sup>139</sup>X. Li and J. Paldus, *J. Chem. Phys.* **131**, 114103 (2009).
- <sup>140</sup>T. Zhang, C. Li, and F. A. Evangelista, *J. Chem. Theory Comput.* **15**, 4399 (2019).
- <sup>141</sup>G. J. R. Aroeira, M. M. Davis, J. M. Turney, and H. F. Schaefer, *J. Chem. Theory Comput.* **17**, 182 (2021).
- <sup>142</sup>P. G. Szalay and R. J. Bartlett, *Chem. Phys. Lett.* **214**, 481 (1993).
- <sup>143</sup>P. G. Szalay and R. J. Bartlett, *J. Chem. Phys.* **103**, 3600 (1995).
- <sup>144</sup>S. A. Kucharski and R. J. Bartlett, *J. Chem. Phys.* **108**, 9221 (1998).
- <sup>145</sup>K. Kowalski and P. Piecuch, *J. Chem. Phys.* **113**, 8490 (2000).
- <sup>146</sup>J. L. Whitten and M. Hackmeyer, *J. Chem. Phys.* **51**, 5584 (1969).
- <sup>147</sup>C. F. Bender and E. R. Davidson, *Phys. Rev.* **183**, 23 (1969).
- <sup>148</sup>B. Huron, J. P. Malrieu, and P. Rancurel, *J. Chem. Phys.* **58**, 5745 (1973).
- <sup>149</sup>R. J. Buenker and S. D. Peyerimhoff, *Theor. Chim. Acta* **35**, 33 (1974).
- <sup>150</sup>J. B. Schriber and F. A. Evangelista, *J. Chem. Phys.* **144**, 161106 (2016).
- <sup>151</sup>J. B. Schriber and F. A. Evangelista, *J. Chem. Theory Comput.* **13**, 5354 (2017).
- <sup>152</sup>N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, *J. Chem. Phys.* **145**, 044112 (2016).
- <sup>153</sup>N. M. Tubman, C. D. Freeman, D. S. Levine, D. Hait, M. Head-Gordon, and K. B. Whaley, *J. Chem. Theory Comput.* **16**, 2139 (2020).
- <sup>154</sup>W. Liu and M. R. Hoffmann, *J. Chem. Theory Comput.* **12**, 1169 (2016).
- <sup>155</sup>N. Zhang, W. Liu, and M. R. Hoffmann, *J. Chem. Theory Comput.* **16**, 2296 (2020).
- <sup>156</sup>A. A. Holmes, N. M. Tubman, and C. J. Umrigar, *J. Chem. Theory Comput.* **12**, 3674 (2016).
- <sup>157</sup>S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, *J. Chem. Theory Comput.* **13**, 1595 (2017).
- <sup>158</sup>J. Li, M. Otten, A. A. Holmes, S. Sharma, and C. J. Umrigar, *J. Chem. Phys.* **149**, 214110 (2018).
- <sup>159</sup>Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, *J. Chem. Phys.* **147**, 034101 (2017).
- <sup>160</sup>Y. Garniron, T. Applencourt, K. Gasperich, A. Benali, A. Ferté, J. Paquier, B. Pradines, R. Assaraf, P. Reinhardt, J. Toulouse, P. Barbaresco, N. Renon, G. David, J.-P. Malrieu, M. Vénil, M. Caffarel, P.-F. Loos, E. Giner, and A. Scemama, *J. Chem. Theory Comput.* **15**, 3591 (2019).
- <sup>161</sup>K. Gururangan, J. E. Deustua, J. Shen, and P. Piecuch, “High-level coupled-cluster energetics by merging moment expansions with selected configuration interaction” (unpublished).