

Efficient multireference perturbation theory without high-order reduced density matrices

Cite as: J. Chem. Phys. **153**, 164120 (2020); <https://doi.org/10.1063/5.0023353>

Submitted: 28 July 2020 . Accepted: 06 October 2020 . Published Online: 29 October 2020

 Nick S. Blunt,  Ankit Mahajan, and Sandeep Sharma

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

[A self-consistent field formulation of excited state mean field theory](#)

The Journal of Chemical Physics **153**, 164108 (2020); <https://doi.org/10.1063/5.0019557>

[Transcorrelated density matrix renormalization group](#)

The Journal of Chemical Physics **153**, 164115 (2020); <https://doi.org/10.1063/5.0028608>

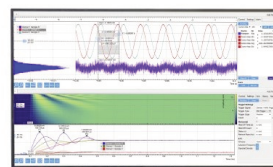
[Models and corrections: Range separation for electronic interaction—Lessons from density functional theory](#)

The Journal of Chemical Physics **153**, 160901 (2020); <https://doi.org/10.1063/5.0028060>



Challenge us.

What are your needs for
periodic signal detection?



Zurich
Instruments

Efficient multireference perturbation theory without high-order reduced density matrices

Cite as: J. Chem. Phys. 153, 164120 (2020); doi: 10.1063/5.0023353

Submitted: 28 July 2020 • Accepted: 6 October 2020 •

Published Online: 29 October 2020



Nick S. Blunt,^{1,a)} Ankit Mahajan,² and Sandeep Sharma^{2,b)}

AFFILIATIONS

¹Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, United Kingdom

²Department of Chemistry, University of Colorado, Boulder, Colorado 80302, 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: nicksblunt@gmail.com

^{b)}Electronic mail: sanshar@gmail.com

ABSTRACT

We present a stochastic approach to perform strongly contracted n -electron valence state perturbation theory (SC-NEVPT), which only requires one- and two-body reduced density matrices, without introducing approximations. We use this method to perform SC-NEVPT2 for complete active space self-consistent field wave functions obtained from selected configuration interaction, although the approach is applicable to a larger class of wave functions, including those from orbital-space variational Monte Carlo. The accuracy of this approach is demonstrated for small test systems, and the scaling is investigated with the number of virtual orbitals and the molecule size. We also find the SC-NEVPT2 energy to be relatively insensitive to the quality of the reference wave function. Finally, the method is applied to the Fe(II)-porphyrin system with a (32e, 29o) active space and to the isomerization of $[\text{Cu}_2\text{O}_2]^{2+}$ in a (28e, 32o) active space.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0023353>

I. INTRODUCTION

In studying electronic structure problems, correlation effects are often separated into strong and dynamic correlation. In some systems, a single determinant is sufficient to provide a qualitative description of a system's electronic structure. However, in strongly correlated systems, this assumption breaks down, and one often has to use a superposition of multiple determinants to describe the reference state. These determinants are obtained by including all (or several) possible occupations within a subset of orbitals known as the active space. One often optimizes the active space orbitals to minimize the energy, which results in a method known as complete active space self-consistent field (CASSCF). The rest of the correlation due to excitation into remaining orbitals is known as dynamic correlation and can be included using a variety of methods including multireference configuration interaction (MRCI),^{1,2} multireference perturbation theory (MRPT),^{3–5} and multireference coupled cluster (MRCC).^{6–11}

In recent years, there have been significant improvements in algorithms for performing (near-exact) CASSCF calculations. Methods including the density matrix renormalization group algorithm

(DMRG),^{12,13} full configuration interaction quantum Monte Carlo (FCIQMC),^{14–17} and selected configuration interaction (SCI)^{18–21} can now be used to solve CASSCF problems accurately for active spaces of 40–50 orbitals and possibly beyond.^{16,17,21–23} However, there still remains the important task of including dynamic correlation. Traditional implementations of MRCI and MRPT require calculating and storing the three- and sometimes four-body reduced density matrices (RDMs) within the active space, which require $\mathcal{O}(n_a^6)$ and $\mathcal{O}(n_a^8)$ storage, respectively. This becomes infeasible for the large active spaces considered above, and separate approaches must be developed.

A variety of methods have been proposed and used to avoid the need for higher-order RDMs. These include the use of cumulant approximations,²⁴ uncontracting terms that require high-order RDMs,²⁵ use of matrix product states,^{26,27} approaches based on FCIQMC (where the high-order RDMs are only sampled),^{28,29} a time-dependent formalism,^{30,31} fully uncontracted formulations of MRPT,³² external contraction,^{33,34} and others.^{35–40}

Recently, we demonstrated that it is possible to perform strongly contracted MRCI (SC-MRCI) and second-order n -electron valence perturbation theory (SC-NEVPT2) without constructing

RDMs but instead using variational Monte Carlo (VMC).⁴¹ In this approach, rather than constructing RDMs and contracting them with integrals, it is possible to directly sample contributions from determinants in the first-order interacting space (FOIS). Although the number of determinants in the FOIS grows exponentially with the active space size, VMC provides a polynomial scaling method to sample them. These stochastic approaches were referred to as SC-MRCI(s) and SC-NEVPT2(s).

In this article, we develop this idea further, focusing specifically on SC-NEVPT2(s). In particular, we present a somewhat different algorithm that is more efficient and avoids the need for a trial wave function in the FOIS, as was required in the original SC-NEVPT2(s) approach. We also extend this to include core orbitals, which were not considered in our original implementation. We go on to provide the analysis of this approach, including scaling with the number of virtual orbitals for the N₂ molecule and with system size for polyacetylene molecules. We also provide examples demonstrating the performance of SC-NEVPT2 when the reference wave function is in error. We then study two much larger systems than considered previously by this method, namely, Fe(II)-porphyrin in a (32e, 29o) active space and [Cu₂O₂]²⁺ in a (28e, 32o) active space, demonstrating that this approach is practical for challenging problems.

II. SC-NEVPT2 OVERVIEW

We begin by recapping the strongly contracted NEVPT2 method^{4,5,42} and defining the notation to be used throughout.

In multireference perturbation theory, one begins by solving the complete active space (CAS) problem, giving a reference wave function $|\phi_m^{(0)}\rangle$,

$$|\phi_m^{(0)}\rangle = \sum_I C_{I,m} |D_I\rangle, \quad (1)$$

where m is the state label and $|D_I\rangle$ are determinants in which all core orbitals (denoted $i, j, \dots$) are occupied and all virtual orbitals (denoted $r, s, \dots$) are unoccupied, while active orbitals (denoted $a, b, \dots$) can take any occupation number.

Assuming that the set of active orbitals is chosen appropriately, $|\phi_m^{(0)}\rangle$ provides a qualitative description of the true wave function. For better accuracy, dynamic correlation must then be included by considering excitations involving core and virtual orbitals. This can be done by second-order perturbation theory, after choosing an appropriate reference Hamiltonian ($\hat{H}_0$). There is no unique way of defining a reference Hamiltonian, in fact, any Hamiltonian that has $|\phi_m^{(0)}\rangle$ as its ground state can be selected. Various reference Hamiltonians have been chosen in the literature, and each leads to a different perturbation theory. In this article, we take the reference Hamiltonian to be the one that defines strongly contracted NEVPT (SC-NEVPT) theory [see Eq. (3)].

In the SC scheme, the uncontracted FOIS is partitioned into subspaces $S_l^{(k)}$. Here, k specifies the change in the number of active space electrons relative to $|\phi_m^{(0)}\rangle$ ($-2 \leq k \leq 2$), while l specifies which non-active orbitals are involved in the excitation. For example, a determinant that contains two unoccupied core orbitals i and j and

a single occupied virtual orbital r belongs to class $S_{ij,r}^{(1)}$. A single perturber state is then assigned to each class $S_l^{(k)}$, defined by

$$|\psi_l^{(k)}\rangle = P_l^{(k)} H |\phi_m^{(0)}\rangle, \quad (2)$$

where $P_l^{(k)}$ is the projector onto the $S_l^{(k)}$ subspace and H is the Hamiltonian operator. This definition ensures that perturber states are orthogonal to each other (but not normalized).

The above perturber states can be further divided into eight types, depending on the number of core and virtual orbitals involved. We refer to these as $v, vv, c, cv, cvv, cc, ccv$, and $ccvv$. For example, we say that a perturber state $|\psi_{ij,r}^{(1)}\rangle$ is of type ccv .

Given the perturber states $|\psi_l^{(k)}\rangle$, the zeroth-order Hamiltonian for SC-NEVPT is defined as

$$H^{(0)} = \sum_m E_m^{(0)} |\phi_m^{(0)}\rangle \langle \phi_m^{(0)}| + \sum_{l,k} E_l^{(k)} |\psi_l^{(k)}\rangle \langle \psi_l^{(k)}|, \quad (3)$$

which leads to the second-order perturbative energy correction,

$$E_m^{(2)} = \sum_{l,k} \frac{N_l^{(k)}}{E_m^{(0)} - E_l^{(k)}}. \quad (4)$$

Here, $N_l^{(k)}$ are the squared norms of the perturbors,

$$N_l^{(k)} = \langle \psi_l^{(k)} | \psi_l^{(k)} \rangle, \quad (5)$$

$E_m^{(0)}$ is the zeroth-order energy for state m , and $E_l^{(k)}$ are the perturber energies. In NEVPT, these perturber energies are defined via the Dyall Hamiltonian, H^D ,

$$E_l^{(k)} = \frac{1}{N_l^{(k)}} \langle \psi_l^{(k)} | H^D | \psi_l^{(k)} \rangle, \quad (6)$$

with

$$H^D = \sum_i^{\text{core}} \epsilon_i a_i^\dagger a_i + \sum_a^{\text{virtual}} \epsilon_a a_a^\dagger a_a + H_{\text{active}}, \quad (7)$$

where H_{active} is the core-averaged Hamiltonian in the active space such that $H^D |\phi_m^{(0)}\rangle = E_m^{(0)} |\phi_m^{(0)}\rangle$.

The primary task is to calculate the second-order energy from Eq. (4). To do so, both the squared norms, $N_l^{(k)}$, and the perturber energies, $E_l^{(k)}$, are required. The exact expressions for $E_l^{(k)}$ and $N_l^{(k)}$ can be obtained in terms of active-space RDMs. However, these include three- and four-body RDMs, whose storage requirements scale as $\mathcal{O}(n_a^6)$ and $\mathcal{O}(n_a^8)$ in the number of active-space orbitals, n_a . Instead, we will take a stochastic approach that avoids the need for higher-order RDMs.

III. STOCHASTIC SC-NEVPT2

The estimation of $E_m^{(2)}$ can be performed in two stages. First, we calculate the squared norms, $N_l^{(k)}$, for all perturbors. In the second step, we calculate the summation in Eq. (4) stochastically by sampling perturbors $|\psi_l^{(k)}\rangle$ with probabilities proportional to $N_l^{(k)}$. For the selected perturber, we estimate the energy $E_l^{(k)}$ and accumulate the contribution toward $E_m^{(2)}$ (this is fully described in Sec. III D).

We begin with some important general points. First, we only wish to avoid the use of three- and four-body RDMs; storing one- and two-body RDMs is always straightforward for current active spaces. We, therefore, use the existing SC-NEVPT2 approach to calculate all instances of $E_l^{(k)}$ and $N_l^{(k)}$, which only require the 1-RDM and 2-RDM. This greatly reduces the sampling task to be performed. Using this rule, in the stochastic approach to be described, we can ignore $ccvv$, cvv , and ccv contributions entirely. For cc , cv , and vv , we need to sample $E_l^{(k)}$ but not $N_l^{(k)}$. For the remaining two sets (c and v), both $E_l^{(k)}$ and $N_l^{(k)}$ must be sampled. This is summarized in Table I.

For the algorithm to be presented, the only computational requirement on $|\phi_m^{(0)}\rangle$ is that overlaps such as $\langle n|\phi_m^{(0)}\rangle$ can be calculated and that the 1-RDM and 2-RDM can be constructed. In this article, we solely take $|\phi_m^{(0)}\rangle$ from SCI. However, this requirement is met by other wave functions such as those that can be optimized in orbital-space VMC. These include Jastrow antisymmetric geminal power wave functions^{43,44} and other symmetry-projected Jastrow mean-field wave functions,⁴⁵ which can be accurate for strong correlation and permit the efficient calculation of $\langle n|\phi_m^{(0)}\rangle$ (though wave functions used in real-space VMC will not).

Whenever generating connections by the application of the Hamiltonian, it should be understood that the heat bath criteria are applied. This was described in the initial presentation of our VMC approach, which we refer to for in-depth description.⁴⁶ This ensures that, for a determinant $|n\rangle$, connections $|p\rangle$ are not generated if $|\langle p|\hat{H}|n\rangle| < \epsilon$ for some small threshold ϵ . For results in this article, we always take $\epsilon = 10^{-8}$ hartree. This typically reduces both the prefactor and scaling of the resulting algorithm, with a negligible effect on the accuracy.

A. The continuous time Monte Carlo algorithm

In the following, it is necessary to sample from probability distributions ρ_n , which take the form

$$\rho_n = \frac{|\langle n|\psi\rangle|^2}{\langle\psi|\psi\rangle}, \quad (8)$$

TABLE I. Table showing which $E_l^{(k)}$ and $N_l^{(k)}$ instances are calculated stochastically and by the traditional deterministic approach. The deterministic approach is taken if only 1- and 2-body RDMs are required, otherwise we use the stochastic approach in order to avoid 3- and 4-body RDMs.

Perturber type	Energies ($E_l^{(k)}$)	Norms ($N_l^{(k)}$)
c	Stochastic	Stochastic
v	Stochastic	Stochastic
cc	Stochastic	Exact
cv	Stochastic	Exact
vv	Stochastic	Exact
ccv	Exact	Exact
cvv	Exact	Exact
$ccvv$	Exact	Exact

where $|\psi\rangle$ is some wave function. Typically, in VMC, this would be sampled by the Metropolis–Hastings algorithm.^{47–49} However, this can be quite inefficient when working in a discrete basis of Slater determinants. Instead, we use the Continuous Time Monte Carlo (CTMC) algorithm,^{50,51} which was introduced to VMC recently.⁴⁶ When applied in other areas, this algorithm is sometimes known as Kinetic Monte Carlo (KMC) or the Bortz–Kalos–Lebowitz (BKL) algorithm. We briefly recap it here:

1. From a determinant $|n\rangle$, calculate $r(p \leftarrow n)$,

$$r(p \leftarrow n) = \frac{|\langle p|\psi\rangle|}{|\langle n|\psi\rangle|}, \quad (9)$$

for all determinants $|p\rangle$ connected to $|n\rangle$ by a single or double excitation (within the relevant space).

2. Calculate the residence time for $|n\rangle$, defined as

$$t_n = \frac{1}{\sum_p r(p \leftarrow n)}. \quad (10)$$

This will define the weight of contributions from $|n\rangle$ in subsequent estimators.

3. Select a new determinant $|p\rangle$ with probability proportional to $r(p \leftarrow n)$.

After a short burn-in period, iterating this procedure will correctly sample ρ_n , provided that t_n are used as weights for contributions to estimators. We denote the total residence time for a random walk by $T = \sum_n t_n$. It is worth pointing out that, in CTMC, all moves are accepted and there are no rejections, but this comes at the added cost of having to evaluate all the overlap ratios in Eq. (9). However, this additional cost is mitigated in our VMC algorithm because these overall ratios are obtained when evaluating the local energy.

B. Previous SC-NEVPT2(s) algorithm

For completeness, we briefly recap the stochastic SC-NEVPT2 approach that we presented previously.⁴¹ This was a trial implementation that excluded excitations involving core orbitals. The energies $E_l^{(k)}$ were estimated by sampling the numerators and denominators,

$$\frac{\langle\psi_l^{(k)}|H^D|\psi_l^{(k)}\rangle}{\langle\psi_s|\psi_s\rangle} = \sum_n \frac{|\langle n|\psi_s\rangle|^2}{\langle\psi_s|\psi_s\rangle} \frac{\langle\psi_l^{(k)}|n\rangle}{\langle\psi_s|n\rangle} \frac{\langle n|H^D|\psi_l^{(k)}\rangle}{\langle n|\psi_s\rangle}, \quad (11)$$

$$\frac{\langle\psi_l^{(k)}|\psi_l^{(k)}\rangle}{\langle\psi_s|\psi_s\rangle} = \sum_n \frac{|\langle n|\psi_s\rangle|^2}{\langle\psi_s|\psi_s\rangle} \frac{|\langle\psi_l^{(k)}|n\rangle|^2}{|\langle\psi_s|n\rangle|^2}, \quad (12)$$

where $|\psi_s\rangle$ is an appropriately chosen sampling wave function,

$$|\psi_s\rangle = \sum_{k,l} c_l^{(k)} |\psi_l^{(k)}\rangle, \quad (13)$$

for some coefficients $c_l^{(k)}$. Here, all quantities are sampled by a single random walk that takes place in $S_0^{(0)}$ and its FOIS, using the CTMC algorithm to sample $\frac{|\langle n|\psi_s\rangle|^2}{\langle\psi_s|\psi_s\rangle}$. This approach was successful for the test systems considered in Ref. 41 but has drawbacks that can be improved upon. First, the approach requires the user to

choose the coefficients $c_l^{(k)}$ in the sampling wave function. A suitable choice is far from obvious, and a suitable prescription can vary widely between systems. The final statistical error varies significantly depending on $|\psi_s\rangle$. In our previous work, we chose $c_l^{(k)}$ randomly with $c_0^{(0)}$ being an order of magnitude larger than other coefficients. This was found to be sufficient for the test systems considered but is far from optimal.

Second, using the CTMC algorithm as described in Sec. III A to sample $\frac{| \langle n | \psi_s \rangle |^2}{\langle \psi_s | \psi_s \rangle}$ requires calculating overlaps $\langle p | \psi_s \rangle$ for all $|p\rangle$ connected to the current walker, $|n\rangle$. In contrast, to calculate the expression Eq. (11) for a given $|n\rangle$ requires

$$\langle n | H^D | \psi_l^{(k)} \rangle = \sum_{p \in S_l^{(k)}} \langle n | H^D | p \rangle \langle p | \psi_l^{(k)} \rangle. \quad (14)$$

Because H^D only acts within the active space, the number of overlaps $\langle p | \psi_l^{(k)} \rangle$ to calculate is much smaller. Thus, the CTMC approach previously used does *not* only use information already available but requires significant extra computation to calculate the additional overlaps.

We instead describe an alternative algorithm that avoids both issues. Here, we first estimate the norms $N_l^{(k)}$ and then use these to sample $S_l^{(k)}$. Each corresponding $E_l^{(k)}$ is then estimated by a random walk entirely within $S_l^{(k)}$.

C. Sampling $N_l^{(k)}$

We take the general case where $|\phi_m^{(0)}\rangle$ may not be normalized. The squared norms can be sampled using the following approach:

$$N_l^{(k)} = \frac{\langle \psi_l^{(k)} | \psi_l^{(k)} \rangle}{\langle \phi_m^{(0)} | \phi_m^{(0)} \rangle} \quad (15)$$

$$= \frac{\langle \phi_m^{(0)} | \hat{H} \hat{P}_l^{(k)} \hat{H} | \phi_m^{(0)} \rangle}{\langle \phi_m^{(0)} | \phi_m^{(0)} \rangle} \quad (16)$$

$$= \sum_{n \in S_0^{(0)}} \frac{| \langle n | \phi_m^{(0)} \rangle |^2}{\langle \phi_m^{(0)} | \phi_m^{(0)} \rangle} \frac{\langle n | \hat{H} \hat{P}_l^{(k)} \hat{H} | \phi_m^{(0)} \rangle}{\langle n | \phi_m^{(0)} \rangle} \quad (17)$$

$$= \left\langle N_l^{(k)} [n] \right\rangle_{\rho_n}. \quad (18)$$

Here, ρ_n is the probability distribution to be sampled by a random walk,

$$\rho_n = \frac{| \langle n | \phi_m^{(0)} \rangle |^2}{\langle \phi_m^{(0)} | \phi_m^{(0)} \rangle}. \quad (19)$$

The determinants selected, $|n\rangle$, are referred to as walkers. We emphasize that this random walk takes places entirely within the CASCI space ($S_0^{(0)}$). The quantity $N_l^{(k)}[n]$ is defined by

$$N_l^{(k)}[n] = \frac{\langle n | \hat{H} \hat{P}_l^{(k)} \hat{H} | \phi_m^{(0)} \rangle}{\langle n | \phi_m^{(0)} \rangle} \quad (20)$$

$$= \frac{\sum_{p \in S_l^{(k)}} \langle n | \hat{H} | p \rangle \sum_{r \in S_0^{(0)}} \langle p | \hat{H} | r \rangle \langle r | \phi_m^{(0)} \rangle}{\langle n | \phi_m^{(0)} \rangle}. \quad (21)$$

$N_l^{(k)}[n]$ is calculated by the following steps: First, generate all determinants $|p\rangle$ in $S_l^{(k)}$ that are connected to $|n\rangle$ (calculating $\langle n | \hat{H} | p \rangle$ for each). Then, for each $|p\rangle$, generate all connected determinants $|r\rangle$ within $S_0^{(0)}$ (calculating $\langle p | \hat{H} | r \rangle$ and $\langle r | \phi_m^{(0)} \rangle$ for each).

In practice, instead of calculating $N_l^{(k)}[n]$ for each $S_l^{(k)}$ separately, we accumulate all instances simultaneously, that is, for each walker $|n\rangle \in S_0^{(0)}$, loop over connected determinants $|p\rangle$ in all $S_l^{(k)}$ is considered, accumulating contributions to $N_l^{(k)}[n]$ for each.

The norm and energy of the zeroth-order wave function are sampled in an analogous way during the same random walk. In this article, we take $|\phi_m^{(0)}\rangle$ from SCI, such that the wave function is normalized by construction, and its energy known. Nonetheless, this step is important in general.

Walker moves within $S_0^{(0)}$ are made using the continuous time Monte Carlo (CTMC) algorithm, described above. Importantly, each $r(p \leftarrow n)$ is already constructed in order to obtain $\langle \phi_m^{(0)} | \phi_m^{(0)} \rangle$ so that the CTMC algorithm can be performed essentially for free.

Note that for every $S_l^{(k)}$ sampled, the quantity $\langle p | \psi_l^{(k)} \rangle = \langle p | \hat{H} | \phi_m^{(0)} \rangle$ is calculated for at least one determinant $|p\rangle$ in $S_l^{(k)}$. We can, therefore, keep a list of determinants that have the largest value of $\langle p | \psi_l^{(k)} \rangle$ for each $S_l^{(k)}$ sector (of the determinants reached). These determinants are used to initialize the walkers when sampling the corresponding $E_l^{(k)}$.

As noted in Table I, we only need to sample norms for perturbors of types c and v . However, we also need to generate initial determinants for cc , cv , and vv . Therefore, there are two parameters that specify the sampling in this step, which we denote N_{norm} and N_{init} . For the first N_{norm} iterations, $N_l^{(k)}$ is only sampled for c and v -type perturbors. We then perform N_{init} iterations in which $N_l^{(k)}$ is sampled for all five perturber types (c , v , cc , cv , and vv). The $N_l^{(k)}$ estimates for cc , cv , and vv from this step are *not* used, as we have access to the exact values. Instead, we use the generated initial determinants when sampling $E_l^{(k)}$ in the next step. These final iterations are more expensive. However, we always take $N_{\text{init}} \ll N_{\text{norm}}$, and typically, $N_{\text{init}} = 50$ is more than sufficient.

D. Sampling $E^{(2)}$ and $E_l^{(k)}$

We next consider the sampling of $E_m^{(2)}$ itself, as defined in Eq. (4). This is done by sampling terms in this summation with a probability proportional to $N_l^{(k)}$,

$$E^{(2)} = \sum_{k,l \neq 0} \frac{1}{E^{(0)} - E_l^{(k)}} N_l^{(k)} \quad (22)$$

$$= \left[\sum_{k',l' \neq 0} N_{l'}^{(k')} \right] \times \sum_{k,l \neq 0} \frac{1}{E^{(0)} - E_l^{(k)}} \cdot \frac{N_l^{(k)}}{\sum_{k',l' \neq 0} N_{l'}^{(k')}} \quad (23)$$

$$= \left[\sum_{k,l \neq 0} N_l^{(k)} \right] \times \left\langle \frac{1}{E^{(0)} - E_l^{(k)}} \right\rangle_{\rho(l,k)}, \quad (24)$$

where $\rho(l,k) = \frac{N_l^{(k)}}{\sum_{k',l' \neq 0} N_{l'}^{(k')}} \cdot$ It is straightforward to sample from $\rho(l,k)$ since all $N_l^{(k)}$ values are stored after the initial stage of the algorithm. We also truncate the summation to only include contributions with $N_l^{(k)} \geq 10^{-8}$ as an efficiency improvement, which we do not find to affect the accuracy.

For each $S_l^{(k)}$ selected, the corresponding $E_l^{(k)}$ must then be estimated. This is achieved by a random walk entirely within $S_l^{(k)}$. Specifically,

$$E_l^{(k)} = \frac{\langle \psi_l^{(k)} | \hat{H}_D | \psi_l^{(k)} \rangle}{\langle \psi_l^{(k)} | \psi_l^{(k)} \rangle} \quad (25)$$

$$= \sum_{n \in S_l^{(k)}} \frac{|\langle n | \psi_l^{(k)} \rangle|^2}{\langle \psi_l^{(k)} | \psi_l^{(k)} \rangle} \frac{\langle n | \hat{H}_D | \psi_l^{(k)} \rangle}{\langle n | \psi_l^{(k)} \rangle} \quad (26)$$

$$= \langle E_L^D[n] \rangle_{\rho_n}, \quad (27)$$

where

$$\rho_n = \frac{|\langle n | \psi_l^{(k)} \rangle|^2}{\langle \psi_l^{(k)} | \psi_l^{(k)} \rangle} \quad (28)$$

and $E_L^D[n]$ is the local energy at $|n\rangle$ with respect to $\hat{H}_D$,

$$E_L^D[n] = \frac{\langle n | \hat{H}_D | \psi_l^{(k)} \rangle}{\langle n | \psi_l^{(k)} \rangle} \quad (29)$$

$$= \frac{\sum_{p \in S_l^{(k)}} \langle n | \hat{H}_D | p \rangle \sum_{r \in S_0^{(0)}} \langle p | \hat{H} | r \rangle \langle r | \phi_m^{(0)} \rangle}{\sum_{r \in S_0^{(0)}} \langle n | \hat{H} | r \rangle \langle r | \phi_m^{(0)} \rangle}. \quad (30)$$

The numerator of this expression is calculated by the following steps: First, generate all connections $|p\rangle$ within $S_l^{(k)}$ (and calculate each $\langle n | \hat{H}_D | p \rangle$). Then, for each $|p\rangle$, generate all connections $|r\rangle$ within $S_0^{(0)}$ (and calculate each $\langle p | \hat{H} | r \rangle$ and $\langle r | \phi_m^{(0)} \rangle$). Similarly, the denominator of this expression is obtained by looping over all connected determinants $|r\rangle$ in $S_0^{(0)}$ and calculating $\langle n | \hat{H} | r \rangle$ and $\langle r | \phi_m^{(0)} \rangle$ for each.

The distribution ρ_n is again sampled using the CTMC algorithm. All required values of $r(p \leftarrow n)$ are obtained when $E_L^D[n]$ is calculated such that this can be performed essentially for free.

There are two parameters that define the sampling in this step, which we denote N_{energy} and $N_{E_l^{(k)}}$. Here, N_{energy} is the number of samples taken from $\rho(l,k)$ (i.e., the number of $E_l^{(k)}$ selected), while $N_{E_l^{(k)}}$ is the number of samples to estimate each $E_l^{(k)}$ selected. However, instead of using a fixed iteration count for all $E_l^{(k)}$, it is often more accurate to use a fixed residence time (see the Appendix). In cases where we use a fixed residence time, we will list both the total residence time used, denoted T , and the average iteration count per $E_l^{(k)}$ estimate.

The final energy is estimated using an average of $1/(E^{(0)} - E_l^{(k)})$ over the selected $S_l^{(k)}$, as in Eq. (24). Because $E_l^{(k)}$ is obtained as a random variable, this is a biased estimator. For a given $E_l^{(k)}$, the expectation value of $1/(E^{(0)} - E_l^{(k)})$ will have an error that will become larger as the statistical error in $E_l^{(k)}$ increases. Because only a small number of samples are used to estimate each $E_l^{(k)}$, this error can become non-negligible for some challenging problems. This is discussed in more detail in the Appendix, where we show that the bias in each $1/(E^{(0)} - E_l^{(k)})$ can be largely corrected by including the following term:

$$E_{\text{bias corr.}} = - \frac{\text{var}[\hat{E}_l^{(k)}]}{(E_m^{(0)} - E_l^{(k)})^3}, \quad (31)$$

where $\text{var}[\hat{E}_l^{(k)}]$ is the variance of the $E_l^{(k)}$ estimate. We include this correction term throughout, unless stated otherwise, with an example given in the Appendix.

E. Parallelism

The above algorithm can be efficiently performed on large-scale computers. In our current implementation, this is done by running the above steps independently on each process. Each process generates its own $N_l^{(k)}$ and $E_l^{(k)}$ estimates and, ultimately, its own $E^{(2)}$ estimate at the end of the simulation. These $E^{(2)}$ values are then averaged to produce the final estimate of the SC-NEVPT2 energy, together with an error estimate. This error estimate is simple to obtain since results from different processes are statistically independent. There is no communication between MPI processes at any point during the simulation.

This approach has very good parallel efficiency. The only cause of non-ideal parallel performance is that processes will take varying times to complete all iterations.

Note that the sampling parameters defined above (N_{norm} , N_{init} , N_{energy} , and $N_{E_l^{(k)}}$) are the number of iterations performed on each process.

F. Scaling

In the following, we denote the number of core, active, and virtual orbitals as n_c , n_a , and n_v , respectively.

In the algorithm presented, a norm estimate $N_l^{(k)}$ is obtained for all (l,k) for which the heat bath criteria are satisfied. However, only a subset of $E_l^{(k)}$ are obtained, as sampled according to the distribution in Eq. (24). We, therefore, consider the scaling to calculate $N_l^{(k)}$ for all (l,k) values and $E_l^{(k)}$ for a constant number of (l,k) samples.

Consider the cost to calculate all $N_l^{(k)}[n]$, for a given $|n\rangle \in S_0^{(0)}$. The expression to be evaluated is given in Eq. (21). First, all determinants $|p\rangle \notin S_0^{(0)}$ connected to $|n\rangle$ are generated. For perturbations of types c , v , cc , cv , and vv , the number of valid $|p\rangle$ scales as $\mathcal{O}(n_a^3 n_c)$, $\mathcal{O}(n_a^3 n_v)$, $\mathcal{O}(n_a^2 n_c^2)$, $\mathcal{O}(n_a^2 n_c n_v)$, and $\mathcal{O}(n_a^2 n_v^2)$, respectively. For each $|p\rangle \in S_l^{(k)}$, the cost to generate all connected $|r\rangle \in S_0^{(0)}$ then scales as $\mathcal{O}(n_a^3)$ for c - and v -type perturbations and as $\mathcal{O}(n_a^2)$ for cc -, cv -, and vv -type perturbations.

TABLE II. The expected scaling to sample $E_l^{(k)}$ or $N_l^{(k)}$ estimates. The scaling presented for $E_l^{(k)}$ is for a fixed (l, k) , while for $N_l^{(k)}$ the scaling is given for all (l, k) of the given perturber type. This assumes that all valid excitations are generated, whereas excitations are actually generated by the heat bath criteria, which are expected to reduce scaling. However, the number of samples required to maintain a constant statistical error will usually increase with system size, increasing the overall scaling. Scaling for real examples is investigated in Sec. IV.

Perturber type	Energies ($E_l^{(k)}$)	Norms ($N_l^{(k)}$)
c	$\mathcal{O}(n_a^7)$	$\mathcal{O}(n_a^6 n_c)$
v	$\mathcal{O}(n_a^7)$	$\mathcal{O}(n_a^6 n_v)$
cc	$\mathcal{O}(n_a^6)$	$\mathcal{O}(n_a^4 n_c^2)$
cv	$\mathcal{O}(n_a^6)$	$\mathcal{O}(n_a^4 n_c n_v)$
vv	$\mathcal{O}(n_a^6)$	$\mathcal{O}(n_a^4 n_v^2)$

$E_l^{(k)}[n]$ is calculated by Eq. (30). For a given $|n\rangle \in S_l^{(k)}$, the cost to generate all connected $|p\rangle \in S_l^{(k)}$ scales as $\mathcal{O}(n_a^4)$ for all perturber types. Then, for each $|p\rangle \in S_l^{(k)}$, the cost to generate all connected $|r\rangle \in S_0^{(0)}$ scales as $\mathcal{O}(n_a^3)$ for c - and v -type perturbors and as $\mathcal{O}(n_a^2)$ for cc -, cv -, and vv -type perturbors.

The overall scaling for each perturber type, obtained from the above arguments, is given in Table II. The true scaling will be somewhat different to this in practice. First, we do not loop over all connected determinants but instead use the heat bath criteria, where connections are not generated if they have a Hamiltonian element below some threshold. This is expected to reduce the overall scaling (however, in this article, we use CASSCF orbitals; because these are delocalized, the potential benefits are more limited). Second, the above only gives the scaling to calculate $E_l^{(k)}[n]$ and $N_l^{(k)}[n]$ for a constant number of samples, $|n\rangle$. In general, the number of samples will increase with system size for a constant statistical error. This increases the overall scaling. It is difficult to write down a general

formula to describe this effect. We instead investigate this through examples in Sec. IV.

IV. RESULTS

In the following, PySCF^{52,53} is used to generate molecular orbitals via CASSCF and molecular integrals for the subsequent SC-NEVPT2(s) calculations. Heat bath CI (HCI) as implemented in the Dice code is used as the CASSCF solver^{21,54,55} and also to generate the zeroth-order wave function $|\phi_m^{(0)}\rangle$ for the SC-NEVPT2(s) step.

To account for burn-in errors, we discard the initial 50 iterations for each CTMC random walk, both for $N_l^{(k)}$ and $E_l^{(k)}$ estimation, unless stated otherwise.

A. Scaling with the number of virtual orbitals: N_2

As a simple first example, we consider N_2 in its ground state at $R = 2.5 a_0$ bond length. The active space is (10e, 8o) with two core orbitals. We then consider calculating the SC-NEVPT2 energy for increasing correlation consistent basis sets from cc-pVDZ (18 virtual orbitals) to aug-cc-pV6Z (368 virtual orbitals).

For the norm-sampling stage, we use parameters $N_{\text{norm}} = 900$ and $N_{\text{init}} = 100$. For the energy sampling stage, we use $N_{\text{energy}} = 10\,000$ and $N_{E_l^{(k)}} = 100$.

The results are presented in Table III. The final two columns compare the stochastic SC-NEVPT2 energies to those calculated with Molpro,⁵⁶ which agree within 1 or 2 statistical error bars. All timings presented are wall times.

The timing and error results from Table III can be used to assess scaling with respect to the number of virtual orbitals. Based on the theoretical scaling in Table II, and for a fixed number of iterations, one would expect the sampling of norms (time t_{norm}) to asymptotically scale with the number of virtual orbitals as $\mathcal{O}(n_v)$. The expected asymptotic scaling to generate initial determinants (time $t_{\text{init.det.}}$) is $\mathcal{O}(n_v^2)$. Sampling a constant number of energies (time t_{energy}) should be independent of n_v .

TABLE III. Scaling of SC-NEVPT2(s) timing and error estimates with basis set size, applied to the ground state of N_2 at $R = 2.5 a_0$. The active space is (10e, 8o). t_{norm} is the time to perform 900 iterations to sample $N_l^{(k)}$ for c - and v -type perturbors. $t_{\text{init.det.}}$ is the time to perform 100 iterations to generate initial determinants. t_{energy} is the time to sample 10 000 values of $E_l^{(k)}$. The final two columns compare the subsequent SC-NEVPT2(s) energy estimates to exact results from Molpro.⁵⁶

Basis	n_v	t_{norm} (s)	$t_{\text{init.det.}}$ (s)	t_{energy} (s)	Statistical error (hartree)	Total energy + 109 (hartree)	
						SC-NEVPT2(s)	Molpro SC-NEVPT2
cc-pVDZ	18	5.103	1.019	234.110	2.0×10^{-4}	−0.1857(2)	−0.185 43
aug-cc-pVDZ	36	10.482	2.454	238.149	2.2×10^{-4}	−0.2025(2)	−0.202 36
cc-pVTZ	50	15.672	3.989	204.581	3.5×10^{-4}	−0.2844(4)	−0.284 98
aug-cc-pVTZ	82	26.841	8.526	225.526	3.6×10^{-4}	−0.2946(4)	−0.294 33
cc-pVQZ	100	30.979	10.756	221.810	5.2×10^{-4}	−0.3452(5)	−0.345 11
aug-cc-pVQZ	150	52.517	21.578	231.923	5.0×10^{-4}	−0.3494(5)	−0.348 80
cc-pV5Z	172	56.710	25.333	224.181	4.1×10^{-4}	−0.3680(4)	−0.367 52
aug-cc-pV5Z	244	94.875	50.289	257.278	3.0×10^{-4}	−0.3706(3)	−0.369 56
cc-pV6Z	270	99.372	55.366	244.260	4.9×10^{-4}	−0.3828(5)	−0.382 85
aug-cc-pV6Z	368	152.935	101.737	271.484	4.6×10^{-4}	−0.3838(5)	−0.384 06

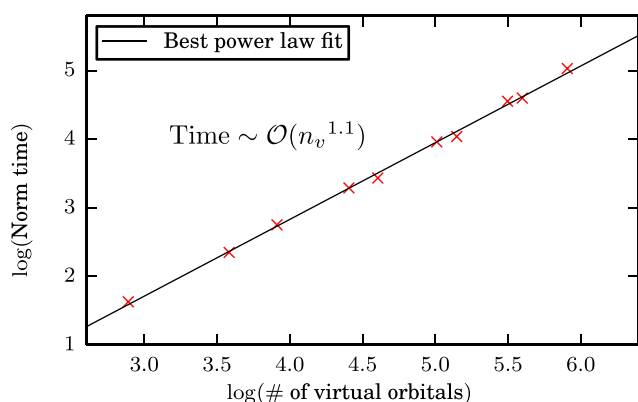


FIG. 1. Scaling of the norm-sampling time (t_{norm}) against the number of virtual orbitals. The system is N_2 at $R = 2.5 a_0$, with a (10e, 8o) active space. A constant number of iterations are performed, $N_{\text{norm}} = 900$. The scaling is found to be $t_{\text{norm}} \sim \mathcal{O}(n_v^{1.1})$.

t_{energy} is seen to be independent of n_v , as expected. Meanwhile, the observed scaling of t_{norm} is $\mathcal{O}(n_v^{1.1})$, while the observed scaling of $t_{\text{init. det.}}$ is $\mathcal{O}(n_v^{1.5})$ in reasonable agreement with the predicted results. The scaling and fit for t_{norm} are shown in Fig. 1.

It is more challenging to reason about how quickly the statistical error should increase. There are two sources of statistical error: first from sampling the norms and second from sampling the energies. We usually observe that the majority of statistical error comes from the energy sampling step, though this will depend on how the simulation parameters are chosen. In the present case, there is a noticeable increase in the statistical error from cc-pVDZ to cc-pVQZ, but interestingly the error becomes somewhat insensitive to n_v beyond this point. There is an error on each of these error estimates, but they are small enough to not affect this conclusion.

B. Scaling with molecule length: Polyacetylene

To consider scaling with overall molecule size, we consider *trans*-polyacetylene molecules with two terminal hydrogen atoms. They take the form $\text{C}_{2n}\text{H}_{2n+2}$. We denote the number of carbon

atoms as N and consider cases from $N = 4$ to $N = 28$. The corresponding number of core, active, and virtual orbitals is given in Table IV.

The orbital basis set is 6-31g. This is not large enough for accurate quantitative results but sufficient for the present scaling study. Similarly, we take a model geometry, where all bond lengths and angles are fixed. Specifically, single C-C bond lengths are 1.45 Å, double C-C bond lengths are 1.34 Å, and C-H bond lengths are 1.08 Å. All angles are set to 120°.

For larger values of N , the CASCI problem becomes infeasible to solve by FCI. Instead, we use selected CI (SCI), specifically the heat bath CI (HCI) method. A constant HCI threshold of $\epsilon = 5 \times 10^{-5}$ hartree is used for each value of N . The number of determinants in the HCI wave function is reported as n_{dets} in Table IV.

The same parameters are used for each simulation: $N_{\text{norm}} = 900$, $N_{\text{init}} = 100$, $N_{\text{energy}} = 1000$, and $N_{E_i^{(k)}} = 100$. Simulations were run on 32 cores on two Intel E5-2650 nodes.

Each of n_c , n_a , and n_v scales linearly with the number of carbon atoms. Therefore, from Table II, the idealized asymptotic scaling for a constant number of iterations is $\mathcal{O}(N^7)$. If we discard the $N = 4$ data point (to better investigate the asymptotic scaling), then the observed scaling for the total time ($t_{\text{norm}} + t_{\text{init. det.}} + t_{\text{energy}}$) is $\mathcal{O}(N^{6.1})$. This lower scaling is reasonable, given that the theoretical scaling does not account for excitations ignored by the heat bath criteria.

There is also an increase in the final statistical error with molecule size. Interestingly, this error decreases from $N = 20$ to $N = 24$; we have checked that this is accurate and not the result of error on the error estimate. However, all other data points follow the expected trend of increasing error.

The statistical error decreases with the number of samples (n_s) as $n_s^{-1/2}$ and so decreases with simulation time (t) as $t^{-1/2}$. Therefore, a sensible measure of overall computational cost is

$$\eta = t \times \sigma^2, \quad (32)$$

where t is the total time and σ is the final error estimate. For the polyacetylene data in Table IV, the values of η are plotted in Fig. 2, which agree well with a linear regression line on this log-log plot. Excluding the first data point ($N = 4$), the overall cost scales roughly as $\mathcal{O}(N^{8.2})$. Although this scaling is steep, it is similar to that of traditional SC-NEVPT2 but with the benefit of not requiring higher-order RDMs. In Sec. IV E, we demonstrate that the method is feasible

TABLE IV. Simulation time and statistical error for SC-NEVPT2(s) simulations performed on polyacetylene, as the number of carbon atoms (N) is increased. n_c , n_a , and n_v give the number of core, active, and virtual orbitals, respectively. A constant number of iterations were performed for each simulation (see the main text for simulation parameters).

No. of C atoms (N)	n_c	n_a	n_v	n_{dets}	t_{norm} (s)	$t_{\text{init. det.}}$ (s)	t_{energy} (s)	Statistical error (hartree)
4	9	4	31	20	1.556	1.165	3.469	1.1×10^{-4}
8	17	8	59	2458	26.469	29.072	90.156	2.2×10^{-4}
12	25	12	87	7.9×10^4	208.94	251.76	865.784	3.0×10^{-4}
16	33	16	115	4.2×10^5	1.134×10^3	1.279×10^3	5.086×10^3	4.8×10^{-4}
20	41	20	143	1.4×10^6	4.694×10^3	4.832×10^3	2.006×10^4	6.8×10^{-4}
24	49	24	171	2.5×10^6	1.981×10^4	1.594×10^4	6.733×10^4	4.7×10^{-4}
28	57	28	199	3.3×10^6	5.615×10^4	4.222×10^4	1.822×10^5	1.1×10^{-3}

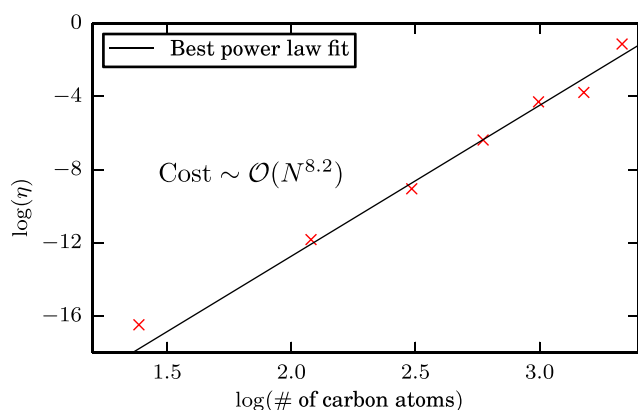


FIG. 2. A measure of computational cost in SC-NEVPT2(s), plotted against the number of carbon atoms (N) in polyacetylene molecules. The cost is $\eta = \sigma \times t^2$, where σ is the statistical error and t is the total simulation time. The cost is seen to scale roughly as $\mathcal{O}(N^{8.2})$.

for active spaces with 32 orbitals. Given the favorable parallel efficiency, we expect active spaces with more than 40 orbitals to be achievable. Nonetheless, we are investigating alternative approaches to reduce this scaling.

C. Effect of error in the reference wave function

It is interesting to investigate how the accuracy of the reference wave function affects the final SC-NEVPT2 energy. This is important in our case since for larger active spaces, we use an approximate HCI wave function as the reference, $|\phi_m^{(0)}\rangle$.

To do this, we have primarily considered the same *trans*-polyacetylene (TPA) system, as studied in Sec. IV B, for the case with

16 carbon atoms, $N = 16$. We performed the SC-NEVPT2(s) procedure using different HCI wave functions, obtained by varying the HCI threshold, ϵ , which controls the accuracy of the wave function. The exact reference is obtained in the small ϵ limit.

For TPA (6-31g) results, the following parameters were used. For the norm sampling step of SC-NEVPT2(s), we take $N_{\text{norm}} = 950$ and $N_{\text{init}} = 50$. For the energy sampling step, N_{energy} is set to 700, except for $\epsilon = 5 \times 10^{-6}$ hartree where $N_{\text{energy}} = 1000$. The total residence time is $T = 1.0$, except for $\epsilon = 5 \times 10^{-6}$ hartree where $T = 1.5$. We use 20 burn-in iterations for norm and energy sampling steps.

To address the concern that results may rely on the very small basis set used, we also obtained results for the same TPA system in the cc-pVDZ basis with two ϵ values. We also performed a similar analysis for the Fe(II)-porphyrin [Fe(P)] system. The system and basis are identical to those fully described in Sec. IV D. The results for $\epsilon = 10^{-5}$ hartree are identical to those presented in Sec. IV D. We then performed an additional calculation with $\epsilon = 3 \times 10^{-5}$ hartree.

The results are given in Table V. For each ϵ value, we state the HCI variational energy, $E^{(0)}$, which is the reference energy in the subsequent SC-NEVPT2 calculation. We also state the Epstein–Nesbet perturbative correction within the CAS (“HCI PT2”), obtained by the semi-stochastic HCI (SHCI) algorithm.⁵⁵ This gives a measure of error in the reference but does not include corrections from the FOIS. We then show the SC-NEVPT2(s) energy estimates $E^{(2)}$ and the final energy estimate, obtained as $E^{(0)} + E^{(2)}$.

The final column can be used to assess the sensitivity of the total SC-NEVPT2 energy to $E^{(0)}$. Interestingly, this total energy shows little variation with ϵ . For TPA (6-31g) with $\epsilon = 5 \times 10^{-4}$ hartree, the HCI variational energy is in error by ~ 38 mhartree, using only 1.2×10^4 determinants in a (16e, 16o) active space. However, the final SC-NEVPT2(s) energy is in error by only ~ 3 mhartree. For $\epsilon = 1 \times 10^{-4}$ hartree, where the reference energy is in error by ~ 10

TABLE V. Results performed for *trans*-polyacetylene (TPA) with 16 carbon atoms ($\text{C}_{16}\text{H}_{18}$) and Fe(II)-porphyrin [Fe(P)] in the 5A_g state. We vary the accuracy of the reference wave function, obtained using the HCI method. We then perform SC-NEVPT2(s) using each resulting reference wave function. The final column shows the variation in the total SC-NEVPT2 energy, which is seen to have only weak dependence on the quality of the reference. Even when the reference energy ($E^{(0)}$) is in error by ~ 38 mhartree, the final SC-NEVPT2 energy is in error by ~ 3 mhartree for TPA (6-31g). We also include the HCI PT2 correction within the CAS, obtained by the semi-stochastic HCI approach.

System	ϵ (hartree)	n_{dets}	Energies (hartree)			
			HCI variational ($E^{(0)}$)	HCI PT2	SC-NEVPT2(s) ($E^{(2)}$)	Total ($E^{(0)} + E^{(2)}$)
TPA (6-31g)	5×10^{-4}	1.2×10^4	−616.2060	−0.0218	−1.2288(4)	−617.4348(4)
	3×10^{-4}	3.7×10^4	−616.2151	−0.0169	−1.2199(5)	−617.4351(5)
	1×10^{-4}	2.4×10^5	−616.2341	−0.0065	−1.2035(5)	−617.4376(5)
	7×10^{-5}	3.3×10^5	−616.2367	−0.0049	−1.2020(6)	−617.4386(6)
	3×10^{-5}	6.3×10^5	−616.2393	−0.0034	−1.1991(5)	−617.4384(5)
	5×10^{-6}	6.4×10^6	−616.2439	−0.0007	−1.1938(5)	−617.4377(5)
TPA (cc-pVDZ)	1×10^{-4}	2.2×10^5	−616.4946	−0.0059	−1.9154(6)	−618.4100(6)
	1×10^{-5}	2.3×10^6	−616.5016	−0.0014	−1.9079(7)	−618.4096(7)
Fe(P)	3×10^{-5}	2.0×10^6	−2245.0225	−0.0061	−3.1708(10)	−2248.1934(10)
	1×10^{-5}	9.3×10^6	−2245.0269	−0.0033	−3.1653(6)	−2248.1922(6)

mhartree, the total SC-NEVPT2 energy is converged to the exact value within statistical error bars. Similarly, TPA (cc-pVDZ) and Fe(P) results show agreement within error bars after varying ϵ .

These results show that the SHCI PT2 energy (which corrects $E^{(0)}$ itself) should not be included in the final energy estimate. Instead, SC-NEVPT2 energies can be estimated simply as $E^{(0)} + E^{(2)}$. Clearly, including the SHCI PT2 correction would give energies in a significant error for the results presented.

The accuracy of $E^{(0)} + E^{(2)}$ can be partially understood because $E^{(2)}$ is formed as a sum of negative quantities, $N_l^{(k)}/(E^{(0)} - E_l^{(k)})$. Therefore, as $E^{(0)}$ becomes less negative (larger ϵ), each contribution in the summation becomes more negative. It is not unreasonable to then expect partial cancellation between errors in $E^{(0)}$ and $E^{(2)}$. Nonetheless, the very accurate nature of cancellation here is perhaps surprising. If this result were general, it would be powerful and extremely useful. However, a general statement on the accuracy of this cancellation cannot be made without more testing, for example, with several different systems and basis sets, which will be a task for future work. However, these are promising initial results, and they justify the HCI wave functions used in Secs. IV D and IV E.

D. Fe(II)-porphyrin

Next, we perform calculations of the Fe(II)-porphyrin [Fe(P)] system. This has been an important benchmark system for multireference methods in recent years, in part due to the difficulty of identifying the spin state ordering.^{17,21,57,58} Experimental results on Fe(P) and related systems have usually found the ground state to be a triplet state, although these results are obtained either from a polar solvent or the crystal phase.^{59–62} Initial theoretical studies have predicted a quintet 5A_g ground state, while a triplet ground state is observed with larger or more careful active space choices.^{21,57,58} Very recently, it has been suggested that the true ground state is a quintet, when geometrical effects are properly considered;⁶³ we do not consider such effects here.

We focus on a (32e, 29o) active space used in early studies by Li Manni *et al.*¹⁷ and subsequently by Smith *et al.*²¹ This active space consists of 20 C $2p_z$, 4 N $2p_z$, and 5 Fe $3d$ orbitals. We then investigate the effect of dynamic correlation through SC-NEVPT2. In particular, we consider the vertical excitation energy, using the same geometry as Smith *et al.*, which is given in the [supplementary material](#). This geometry was originally described by Groenhof *et al.*,⁶⁴ was optimized for the triplet state, and was also used in a DMRG investigation of this system.⁶⁵ At this fixed geometry, previous results suggest that the ground state is a triplet; for example, this was found to be the case with a larger (44e, 44o)²¹ active space. Lee and co-workers also studied this system recently,⁶⁶ giving a useful summary of recent results and using auxiliary-field quantum Monte Carlo (AFQMC) to confirm the triplet ground state. However, for this (32e, 29o) active space, and at the CASSCF level of theory, a 5A_g ground state is observed. It is interesting and valuable to investigate to what extent SC-NEVPT2 can correct this situation.

Although Fe(P) has D_{4h} symmetry, we use D_{2h} instead. Using D_{2h} symmetry labels, we calculate the lowest energy states in both

the 5A_g and $^3B_{1g}$ sectors. Note that the irreducible representation B_{1g} of D_{2h} corresponds to A_{2g} and B_{2g} in D_{4h} .

The basis set is cc-pVDZ. We use the same CASSCF orbitals optimized by Smith *et al.* for their CASSCF study of this system, where HCI was used as the solver. The reference wave function in SC-NEVPT2(s) was also obtained with HCI, using a final threshold of $\epsilon = 10^{-5}$ hartree, which resulted in a wave function of $\sim 10^7$ determinants for both states. For SC-NEVPT2(s) simulations, parameters used by each process were: for the 5A_g state, $N_{\text{norm}} = 950$, $N_{\text{init}} = 50$, $N_{\text{energy}} = 1500$, and $T = 0.4$ (giving $N_{E_l^{(k)}} \approx 72$ on average), and for the $^3B_{1g}$ state, $N_{\text{norm}} = 900$, $N_{\text{init}} = 100$, $N_{\text{energy}} = 1260$, and $T = 0.4$ (giving $N_{E_l^{(k)}} \approx 104$ on average), performed with 320 MPI processes for both states. The following orbitals were frozen in the SC-NEVPT2(s) calculation: 20 C $1s$, 20 N $1s$, and $1-3s$, $2-3p$ on the Fe atom, 33 orbitals in total.

The results are presented in Table VI. Using CASSCF only, the 5A_g state is lower in energy than the $^3B_{1g}$ state by ~ 31 mhartree. Including the SC-NEVPT2 correction, it is seen that the quintet state remains the ground state; however, the energy gap is lowered by approximately 19 mhartree, suggesting an improved result overall.

Note that the CASSCF energy is in error by approximately +5 mhartree for the 5A_g state and by $\sim +9$ mhartree for the $^3B_{1g}$ state due to the finite value of ϵ used in HCI, although correcting for this does not change our conclusion significantly. It would be simple to improve this by using a smaller value of ϵ , which only has a small effect on the SC-NEVPT2(s) simulation time. This is because coefficients in the reference wave function are obtained by a hash table lookup, the time for which has very weak scaling with the number of determinants.

Our results show that including dynamic correlation through SC-NEVPT2 does noticeably improve the predicted energy gap in this system, but that the expected ordering only occurs with a larger active space. In particular, including the set of 5 Fe $4d$ orbitals together with 10 σ bonds between Fe and N atoms (1 Fe $4p_x$, 1 Fe $4p_y$, 4 N $2p_x$, and 4 N $2p_y$) results in a (44e, 44o) active space,^{21,65} which gives a triplet ground state. Li Manni and Alavi have also studied a separate model of Fe(P), where $C_\beta H$ groups are replaced by hydrogen atoms. With this, they also predict a triplet ground state with a more compact (32e, 34o) active space, which also includes the Fe $4d$ orbitals, and part of the Fe–N σ manifold.⁵⁷ Combined, these results highlight the importance of appropriately choosing the active space in such systems.

TABLE VI. Energies for two low-lying states of Fe(II)-porphyrin, obtained with CASSCF and SC-NEVPT2(s), using a common geometry for both states. The (32e, 29o) active space of Li Manni *et al.*¹⁷ was used. Irreducible representation labels here refer to the D_{2h} point group, which was used for all calculations.

State	Energies (hartree)		
	CASSCF	SC-NEVPT2(s)	Total
5A_g	−2245.0269	−3.1653(6)	−2248.1922(6)
$^3B_{1g}$	−2244.9957	−3.1844(7)	−2248.1800(7)
ΔE	0.0312	−0.0190(9)	0.0122(9)

E. $[\text{Cu}_2\text{O}_2]^{2+}$

As a final example, we consider the $[\text{Cu}_2\text{O}_2]^{2+}$ molecule. In particular, we study the isomerization between bis(μ -oxo) and μ - η^2 : η^2 -peroxo isomers. The model and process, in particular, when combined with appropriate ligands, have an important role as an active site for O_2 activation by enzymes such as tyrosinase. Given the presence of transition metals, it is expected that the treatment of static correlation may be important, and it has further been suggested that a balanced treatment of static and dynamic is required for accurate results. Moreover, existing benchmarks are available from previous computational studies,^{67,68} making this a sensible test system.

We describe the isomerization process using the same geometries of Cramer *et al.*⁶⁷ In this, the Cu–Cu distance is equal to $2.8 + 0.8F$ Å, while the O–O distance is equal to $2.3 - 0.9F$ Å. Here, F is a parameter, which varies from 0 to 1. $F = 0$ indicates the bis(μ -oxo) geometry, and $F = 1$ indicates the μ - η^2 : η^2 -peroxo geometry.

We use the ANO-RCC-VQZP basis set,^{69,70} which corresponds to Cu: [21s15p10d6f4g2h/7s6p4d3f2g1h] and O: [14s9p4d3f2g/4s3p2d2f1g] contractions. This is slightly different to the basis used in other studies such as that by Yanai *et al.*⁶⁸

We take the same (28e, 32o) active space of Yanai *et al.*, consisting of all Cu 3d and 4d orbitals and all O 2p and 3p orbitals.

CASSCF orbitals are obtained with HCI using a final threshold of $\epsilon = 10^{-4}$ hartree. We then use a tighter threshold of $\epsilon = 2 \times 10^{-5}$ hartree to generate the reference wave function for SC-NEVPT2(s). This results in HCI wave functions with between 2.0×10^7 and 2.6×10^7 determinants, depending on F . We then perform SC-NEVPT2(s), using norm parameters $N_{\text{norm}} = 450$ and $N_{\text{init}} = 50$. The number of energy samples, N_{energy} , is between 2000 and 2100, and simulations were run with either 320 or 360 processes, depending on the value of F . The total residence time T was set to 0.4, which gave $N_{E_i^{(k)}}$ between 50 and 55 on average (in addition to 50 burn-in iterations).

The results are given in Table VII and plotted in Fig. 3. We also include results from previous studies for comparison. In particular, CAS(16,14), CR-CCSD(TQ), and CR-CCSD(TQ)_L results were

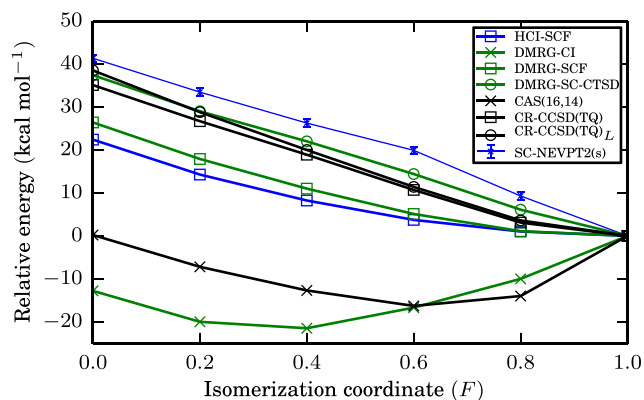


FIG. 3. Energies for the isomerization of $[\text{Cu}_2\text{O}_2]^{2+}$, relative to the μ - η^2 : η^2 -peroxo isomer ($F = 1.0$). Data plotted are the same as in Table VII. Results plotted in black are from Ref. 67. Results plotted in green are from Ref. 68.

taken from the study of Cramer *et al.*,⁶⁷ and DMRG-CI, DMRG-SCF, and DMRG-SC-CTSD results were taken from the study of Yanai *et al.*⁶⁸ Our CASSCF results, obtained using HCI as a solver, are labeled “HCI-SCF.” It is known to be difficult to obtain the correct isomerization profile for this system. Too small an active space leads to an unphysical minimum. HCI-SCF results show that the more substantial (28e, 32o) active space removes this minimum, as previously found by Yanai using DMRG-SCF. More accurate results are obtained when dynamical correlation is included. Our SC-NEVPT2(s) results are approximately in agreement with existing results. We find slightly larger relative energies than previous results. However, we use a larger basis set, so it is perhaps expected that results will not be identical. Cramer *et al.* also use a pseudopotential for Cu atoms, while we freeze core electrons. Overall, these results show reasonable agreement and demonstrate the usefulness of this approach for a significant active space.

V. CONCLUSION

In this work, we have developed a stochastic approach to performing strongly contracted NEVPT2. This method reproduces exact SC-NEVPT2 energies within statistical error bars but avoids the prohibitive cost of constructing and storing three- and four-body RDMs.

The method has low scaling with the number of virtual orbitals, n_v . The cost to sample a fixed number of perturber energies, $E_i^{(k)}$, is independent of n_v , while the increase in associated statistical error is low for small basis sets, plateauing off for larger basis sets. The scaling with the number of active space orbitals is more restrictive. In particular, we investigated the scaling of the overall computational cost with molecular size, N , for polyacetylene molecules. In this case, the number of core, active, and virtual orbitals increases linearly with N , and the total cost (after accounting for the increase in statistical error) was found to scale roughly as $\mathcal{O}(N^{8.2})$.

We also investigated the sensitivity of the final SC-NEVPT2 energy to a reference wave function of varying accuracy. Interestingly, we found final energies to remain accurate, with relatively weak dependence on the quality of the reference energy. If this result

TABLE VII. Energies (in kcal mol⁻¹) from various methods, including SC-NEVPT2(s), for the isomerization of $[\text{Cu}_2\text{O}_2]^{2+}$ between bis(μ -oxo) and μ - η^2 : η^2 -peroxo isomers. Energies are relative to the μ - η^2 : η^2 -peroxo isomer ($F = 1.0$). Note that we use a different basis set to these two studies.

Method	F					
	0	0.2	0.4	0.6	0.8	1
HCI-SCF	22.4	14.3	8.2	3.7	1.0	0
SC-NEVPT2(s)	41.3(8)	33.5(9)	26.3(9)	19.9(8)	9.3(9)	0
CAS(16,14) ^a	0.2	-7.2	-12.7	-16.3	-14.0	0
CR-CCSD(TQ) ^a	35.1	26.7	18.9	10.7	3.1	0
CR-CCSD(TQ) _L ^a	38.5	28.8	20.0	11.4	3.6	0
DMRG-CI ^b	-12.8	-20.9	-21.5	-16.7	-10.0	0
DMRG-SCF ^b	26.4	17.9	11.0	5.1	1.1	0
DMRG-SC-CTSD ^b	37.4	29.0	22.0	14.4	6.1	0

^aResults from Ref. 67.

^bResults from Ref. 68.

were general, then it would be very powerful. We intend to study this for further systems to investigate this possibility.

The method was applied to example systems where multi-reference behavior is expected to be important: Fe(II)-porphyrin with a (32e, 29o) active space and $[\text{Cu}_2\text{O}_2]^{2+}$ with a (28e, 32o) active space. The method was successfully applied to these large active spaces, raising the possibility of obtaining SC-NEVPT2 results, without approximations, in larger active spaces than previously considered. These calculations were performed with moderate computer resources. However, the approach has good parallel efficiency such that it could be used in a straightforward manner on much larger parallel computers, as have been used in many QMC studies previously.

There are several areas in which this method could be developed. First, it will be important to develop SC-NEVPT2(s) to work with other types of wave functions, in particular orbital-space VMC wave functions.⁴⁵ Such wave functions can be well suited to strong correlation and with favorable scaling.^{43,44,71} Because only wave function overlaps ($\langle n|\phi_m^{(0)}\rangle$) and 1- and 2-RDMs are needed, this should be a straightforward task. Our code already supports the optimization of such wave functions, including the calculation of the required overlaps and RDMs. Second, we are keen to investigate approaches to reduce the scaling, in particular, with respect to the active space size. With these developments, we hope that this may be a robust method to perform NEVPT2 with active spaces of 40–50 orbitals, which we believe would be valuable in the general task of performing strongly correlated electronic structure calculations.

SUPPLEMENTARY MATERIAL

The [supplementary material](#) includes the geometry of the Fe(II)-porphyrin model studied in this article. This geometry was taken from Ref. 64. The geometries for all other systems are stated in the article.

ACKNOWLEDGMENTS

N.S.B. is grateful to St John's College, Cambridge for funding and supporting this work through a Research Fellowship. S.S. and A.M. were supported by NSF through Grant No. CHE-1800584. S.S. was also partly supported through the Sloan research fellowship. This study made use of the CSD3 Peta4-Skylake CPU cluster at the University of Cambridge and the Summit supercomputer at CU Boulder.

APPENDIX: POTENTIAL BIASES

Because a large number of energies $E_l^{(k)}$ must be sampled, each with its own independent random walk, only a limited number of samples can be used to estimate each $E_l^{(k)}$. This is different to the typical case in VMC, where a single energy is to be estimated by a long random walk (typically by the Metropolis algorithm). Statistical biases in QMC will become larger as the number of samples becomes smaller. Therefore, for the small number of samples used, there are some potential biases to consider carefully for the algorithm presented.

1. Burn-in

Each random walk with the CTMC algorithm has a burn-in period. In practice, we have found that results are essentially identical regardless of whether burn-in iterations are discarded or not, suggesting this to be a negligible effect here. Nonetheless, it is sensible to account for this possibility where affordable. We, therefore, typically discard the first 50 iterations for each random walk, both in $S_0^{(0)}$ (for $N_l^{(k)}$ estimation) and in each $S_l^{(k)}$ sampled (for $E_l^{(k)}$ estimation).

2. CTMC estimates of $E_l^{(k)}$

Some care is required in using the CTMC algorithm. In CTMC, a sample from a given determinant $|n\rangle$ is weighted by a corresponding residence time, defined as $t_n = \frac{1}{\sum_p r(p \leftarrow n)}$. The final point estimate of $E_l^{(k)}$ is obtained by

$$\hat{E}_l^{(k)} = \frac{\sum_n t_n E_l^D[n]}{\sum_n t_n}. \quad (\text{A1})$$

Using a constant number of iterations for each $E_l^{(k)}$ leads to small statistical bias, which becomes noticeable for very large systems. Instead, each $E_l^{(k)}$ should be estimated with a constant total residence time, $T = \sum_n t_n$. We, therefore, run CTMC random walks until some fixed threshold time is reached, at which point the walk is ended. This is found to resolve all such issues with biases in $E_l^{(k)}$ estimates.

3. Bias in $(E_m^{(0)} - E_l^{(k)})^{-1}$ estimator

Contributions to $E_m^{(2)}$ each take the form $(E_m^{(0)} - E_l^{(k)})^{-1}$, where each $E_l^{(k)}$ is stochastically sampled. Even if the estimator for $E_l^{(k)}$ is unbiased, the final result will be biased because $E[\frac{1}{X}] \neq \frac{1}{E[X]}$. Estimators of this type are very common in QMC, and associated biases are typically negligible. In the current case, however, the bias is larger because the number of samples used to estimate each $E_l^{(k)}$ is very small (~50–100) for the calculations presented in this work.

To see the issue more clearly, we can consider a Taylor expansion of $(E_m^{(0)} - \hat{E}_l^{(k)})^{-1}$, where $\hat{E}_l^{(k)}$ is a point estimate of $E_l^{(k)}$. We may write $\hat{E}_l^{(k)} = E_l^{(k)} + \delta$, where δ denotes the error. Assuming that $\hat{E}_l^{(k)}$ is unbiased, we have that $E[\delta] = 0$. One can then write

$$\frac{1}{E_m^{(0)} - \hat{E}_l^{(k)}} = \frac{1}{E_m^{(0)} - E_l^{(k)} - \delta} \quad (\text{A2})$$

$$= \frac{1}{(E_m^{(0)} - E_l^{(k)}) \left[1 - \frac{\delta}{E_m^{(0)} - E_l^{(k)}} \right]} \quad (\text{A3})$$

$$= \frac{1}{E_m^{(0)} - E_l^{(k)}} \left[1 + \frac{\delta}{E_m^{(0)} - E_l^{(k)}} + \frac{\delta^2}{(E_m^{(0)} - E_l^{(k)})^2} + \mathcal{O}(\delta^3) \right]. \quad (\text{A4})$$

We can use this to look at the expected value of $(E_m^{(0)} - \hat{E}_l^{(k)})^{-1}$,

$$\mathbb{E}\left[\frac{1}{E_m^{(0)} - \hat{E}_l^{(k)}}\right] = \frac{1}{E_m^{(0)} - E_l^{(k)}} \left[1 + \frac{\mathbb{E}[\delta^2]}{(E_m^{(0)} - E_l^{(k)})^2} + \mathcal{O}(\delta^3)\right] \quad (\text{A5})$$

$$= \frac{1}{E_m^{(0)} - E_l^{(k)}} + \frac{\text{var}[\hat{E}_l^{(k)}]}{(E_m^{(0)} - E_l^{(k)})^3} + \mathcal{O}(\delta^3). \quad (\text{A6})$$

Therefore, it can be seen that the bias will increase as the energy difference $E_m^{(0)} - E_l^{(k)}$ becomes smaller and as the estimate of $E_l^{(k)}$ becomes more noisy.

The above equation gives an expression to correct much of the bias,

$$E_{\text{bias corr.}} = -\frac{\text{var}[\hat{E}_l^{(k)}]}{(E_m^{(0)} - E_l^{(k)})^3}. \quad (\text{A7})$$

Using this expression requires an estimate of the variance of $\hat{E}_l^{(k)}$. If the Metropolis algorithm were used, the standard estimator for the variance of the mean would be used,

$$\hat{\sigma}_{E_l^{(k)}}^2 = \frac{1}{N_s(N_s - 1)} \sum_n (E_L^D[n] - \overline{E_L^D})^2, \quad (\text{A8})$$

where N_s is the number of samples and $\overline{E_L^D}$ is the sample mean. Instead, we use the CTMC algorithm, where the estimator for $E_l^{(k)}$ is formed as a weighted sum, as in Eq. (A1). An estimator for the variance of a weighted sum is more complicated, and there is no generally accepted formula for all applications. We have tested several estimators and found that the following formula^{72,73} is very accurate for our case, which we, therefore, use:

$$\hat{\sigma}_{E_l^{(k)}}^2 = \frac{1}{T} \frac{N_s}{(N_s - 1)} \left[\sum_n (t_n E_L^D[n] - T \overline{E_L^D})^2 \right. \quad (\text{A9})$$

$$\left. - 2 \overline{E_L^D} \sum_n (t_n - T)(t_n E_L^D[n] - T \overline{E_L^D}) \right. \quad (\text{A10})$$

$$\left. + \overline{E_L^D}^2 \sum_n (t_n - T)^2 \right]. \quad (\text{A11})$$

Here, $T = \sum_n t_n$ is the total residence time and t_n act as weights in the estimator for $E_l^{(k)}$, as in Eq. (A1). $E_L^D[n]$ is the local energy with respect to the Dyll Hamiltonian, as in Eq. (29). In addition, samples $E_L^D[n]$ are serially correlated, and we account for this by using an automated reblocking procedure.⁷⁴

4. Example: N₂ cc-pVDZ

As a simple example to demonstrate these concepts, in particular, the estimation of $\sigma_{E_l^{(k)}}^2$ and the bias correction term, we consider N₂ in a cc-pVDZ basis set. This is the same example considered in Sec. IV A, using a (10e, 8o) active space and two core orbitals.

We consider the estimation of a single perturber energy, $E_l^{(k)}$, of type $\nu\nu$, involving the two virtual orbitals that are lowest in energy. For this small example, it is possible to enumerate all determinants in $S_l^{(k)}$ and calculate the exact $E_l^{(k)}$. By repeating the stochastic estimation of $E_l^{(k)}$ a large number of times, we can investigate the above

effects. In particular, we repeat this estimation of $E_l^{(k)}$ 100 000 times so that we can accurately construct the distribution function and investigate the true variance and bias.

For the perturber in question, the exact result is $E_l^{(k)} = -105.25896$ hartree. Performing the CTMC estimation of $E_l^{(k)}$, exactly as in the SC-NEVPT2(s) algorithm, and then averaging over the 100 000 repeated estimates give $E_l^{(k)} = -105.25882(15)$ hartree so that the method is unbiased within error bars. An accurate estimate of the variance (obtained directly from the constructed probability distribution) is $\text{Var}[\hat{E}_l^{(k)}] = 0.00229$ hartree², while the estimate from Eq. (A11) is $\sigma_{E_l^{(k)}}^2 = 0.00233$ hartree².

Similarly, the difference between the exact and estimated values of $(E_m^{(0)} - E_l^{(k)})^{-1}$ is $3.3(11) \times 10^{-5}$ hartree⁻¹, indicating the possibility of a small bias. Including the above bias correction changes this discrepancy to $-1.1(11) \times 10^{-5}$ hartree⁻¹, suggesting an improvement. In this case, the correction is extremely small so could be ignored. For non-trivial problems, this correction needs more careful consideration. For the [Cu₂O₂]²⁺ examples in Sec. IV E, the bias correction in Eq. (A7) is of size ≈ 0.6 mhartree for each value of F . We, therefore, include this correction term in all results presented in this article.

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

- H. J. Werner and P. J. Knowles, *J. Chem. Phys.* **89**, 5803 (1988).
- P. J. Knowles and H.-J. Werner, *Theor. Chim. Acta* **84**, 95 (1992).
- K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, *J. Phys. Chem.* **94**, 5483 (1990).
- C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu, *J. Chem. Phys.* **114**, 10252 (2001).
- C. Angeli, R. Cimiraglia, and J.-P. Malrieu, *J. Chem. Phys.* **117**, 9138 (2002).
- R. J. Bartlett and M. Musiał, *Rev. Mod. Phys.* **79**, 291 (2007).
- F. A. Evangelista, *J. Chem. Phys.* **149**, 030901 (2018).
- U. S. Mahapatra, B. Datta, and D. Mukherjee, *J. Chem. Phys.* **110**, 6171 (1999).
- J. Meller, J. P. Malrieu, and R. Caballol, *J. Chem. Phys.* **104**, 4068 (1996).
- J. Mášik and I. Hubač, *Adv. Quantum Chem.* **31**, 75 (1998).
- E. Giner, G. David, A. Scemama, and J. P. Malrieu, *J. Chem. Phys.* **144**, 064101 (2016).
- S. R. White, *Phys. Rev. Lett.* **69**, 2863 (1992).
- G. K.-L. Chan and M. Head-Gordon, *J. Chem. Phys.* **116**, 4462 (2002).
- G. H. Booth, A. J. W. Thom, and A. Alavi, *J. Chem. Phys.* **131**, 054106 (2009).
- D. Cleland, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **132**, 041103 (2010).
- R. E. Thomas, Q. Sun, A. Alavi, and G. H. Booth, *J. Chem. Theory Comput.* **11**, 5316 (2015).
- G. Li Manni, S. D. Smart, and A. Alavi, *J. Chem. Theory Comput.* **12**, 1245 (2016).
- B. Huron, J. P. Malrieu, and P. Rancurel, *J. Chem. Phys.* **58**, 5745 (1973).
- S. Evangelisti, J.-P. Daudey, and J.-P. Malrieu, *Chem. Phys.* **75**, 91 (1983).
- Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, *J. Chem. Phys.* **147**, 034101 (2017).
- J. E. T. Smith, B. Mussard, A. A. Holmes, and S. Sharma, *J. Chem. Theory Comput.* **13**, 5468 (2017).
- D. Zgid and M. Nooijen, *J. Chem. Phys.* **128**, 144116 (2008).

- ²³D. Ghosh, J. Hachmann, T. Yanai, and G. K.-L. Chan, *J. Chem. Phys.* **128**, 144117 (2008).
- ²⁴D. Zgid, D. Ghosh, E. Neuscamman, and G. K.-L. Chan, *J. Chem. Phys.* **130**, 194107 (2009).
- ²⁵P. Celani and H.-J. Werner, *J. Chem. Phys.* **112**, 5546 (2000).
- ²⁶S. Sharma and G. K.-L. Chan, *J. Chem. Phys.* **141**, 111101 (2014).
- ²⁷S. Sharma, G. Knizia, S. Guo, and A. Alavi, *J. Chem. Theory Comput.* **13**, 488 (2017).
- ²⁸R. J. Anderson, T. Shiozaki, and G. H. Booth, *J. Chem. Phys.* **152**, 054101 (2020).
- ²⁹J. J. Halson, R. J. Anderson, and G. H. Booth, "Improved stochastic multireference perturbation theory for correlated systems with large active spaces," *Mol. Phys.* (published online).
- ³⁰A. Y. Sokolov and G. K.-L. Chan, *J. Chem. Phys.* **144**, 064102 (2016).
- ³¹A. Y. Sokolov, S. Guo, E. Ronca, and G. K.-L. Chan, *J. Chem. Phys.* **146**, 244102 (2017).
- ³²E. Giner, C. Angeli, Y. Garniron, A. Scemama, and J.-P. Malrieu, *J. Chem. Phys.* **146**, 224108 (2017).
- ³³P. E. M. Siegbahn, *Int. J. Quantum Chem.* **23**, 1869 (1983).
- ³⁴Z. Luo, Y. Ma, X. Wang, and H. Ma, *J. Chem. Theory Comput.* **14**, 4747 (2018).
- ³⁵L. Gagliardi, D. G. Truhlar, G. Li Manni, R. K. Carlson, C. E. Hoyer, and J. L. Bao, *Acc. Chem. Res.* **50**, 66 (2017).
- ³⁶E. Pastorczak, M. Hapka, L. Veis, and K. Pernal, *J. Phys. Chem. Lett.* **10**, 4668 (2019).
- ³⁷E. Giner, B. Pradines, A. Ferté, R. Assaraf, A. Savin, and J. Toulouse, *J. Chem. Phys.* **149**, 194301 (2018).
- ³⁸C. Li and F. A. Evangelista, *Annu. Rev. Phys. Chem.* **70**, 245 (2019).
- ³⁹J. E. Deustua, I. Magoulas, J. Shen, and P. Piecuch, *J. Chem. Phys.* **149**, 151101 (2018).
- ⁴⁰F. M. Faulstich, M. Máté, A. Laestadius, M. A. Csirik, L. Veis, A. Antalík, J. Brabec, R. Schneider, J. Pittner, S. Kvaal, and Ö. Legeza, *J. Chem. Theory Comput.* **15**, 2206 (2019).
- ⁴¹A. Mahajan, N. S. Blunt, I. Sabzevari, and S. Sharma, *J. Chem. Phys.* **151**, 211102 (2019).
- ⁴²C. Angeli, R. Cimiraglia, and J.-P. Malrieu, *Chem. Phys. Lett.* **350**, 297 (2001).
- ⁴³E. Neuscamman, *Phys. Rev. Lett.* **109**, 203001 (2012).
- ⁴⁴E. Neuscamman, *J. Chem. Phys.* **139**, 181101 (2013).
- ⁴⁵A. Mahajan and S. Sharma, *J. Phys. Chem. A* **123**, 3911 (2019).
- ⁴⁶I. Sabzevari and S. Sharma, *J. Chem. Theory Comput.* **14**, 6276 (2018).
- ⁴⁷N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- ⁴⁸D. Ceperley, G. V. Chester, and M. H. Kalos, *Phys. Rev. B* **16**, 3081 (1977).
- ⁴⁹W. M. C. Foulkes, L. Mitás, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
- ⁵⁰A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, *J. Comput. Phys.* **17**, 10 (1975).
- ⁵¹D. T. Gillespie, *J. Comput. Phys.* **22**, 403 (1976).
- ⁵²Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. L. Chan, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, e1340 (2018).
- ⁵³Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, *J. Chem. Phys.* **153**, 024109 (2020).
- ⁵⁴A. A. Holmes, N. M. Tubman, and C. J. Umrigar, *J. Chem. Theory Comput.* **12**, 3674 (2016).
- ⁵⁵S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, *J. Chem. Theory Comput.* **13**, 1595 (2017).
- ⁵⁶H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz *et al.*, Molpro, version 2015.1, a package of *ab initio* programs, 2015, <http://www.molpro.net>.
- ⁵⁷G. Li Manni and A. Alavi, *J. Phys. Chem. A* **122**, 4935 (2018).
- ⁵⁸G. Li Manni, D. Kats, D. P. Tew, and A. Alavi, *J. Chem. Theory Comput.* **15**, 1492 (2019).
- ⁵⁹T. Kitagawa and J. Teraoka, *Chem. Phys. Lett.* **63**, 443 (1979).
- ⁶⁰J. Mispelter, M. Momenteau, and J. M. Lhoste, *J. Chem. Phys.* **72**, 1003 (1980).
- ⁶¹M. Evangelisti, J. Bartolomé, L. J. de Jongh, and G. Filoti, *Phys. Rev. B* **66**, 144410 (2002).
- ⁶²J. Bartolomé, F. Bartolomé, L. M. García, G. Filoti, T. Gredig, C. N. Colesniuc, I. K. Schuller, and J. C. Cezar, *Phys. Rev. B* **81**, 195405 (2010).
- ⁶³A. Antalík *et al.*, *Phys. Chem. Chem. Phys.* **22**, 17033 (2020).
- ⁶⁴A. R. Groenhof, M. Swart, A. W. Ehlers, and K. Lammertsma, *J. Phys. Chem. A* **109**, 3411 (2005).
- ⁶⁵R. Olivares-Amaya, W. Hu, N. Nakatani, S. Sharma, J. Yang, and G. K.-L. Chan, *J. Chem. Phys.* **142**, 034102 (2015).
- ⁶⁶J. Lee, F. D. Malone, and M. A. Morales, *J. Chem. Theory Comput.* **16**, 3019 (2020).
- ⁶⁷C. J. Cramer, M. Włoch, P. Piecuch, C. Puzzarini, and L. Gagliardi, *J. Phys. Chem. A* **110**, 1991 (2006).
- ⁶⁸T. Yanai, Y. Kurashige, E. Neuscamman, and G. K.-L. Chan, *J. Chem. Phys.* **132**, 024105 (2010).
- ⁶⁹B. O. Roos, R. Lindh, P.-Å. Malmqvist, V. Veryazov, and P.-O. Widmark, *J. Phys. Chem. A* **108**, 2851 (2004).
- ⁷⁰B. O. Roos, R. Lindh, P.-Å. Malmqvist, V. Veryazov, and P.-O. Widmark, *J. Phys. Chem. A* **109**, 6575 (2005).
- ⁷¹M. Casula, C. Attaccalite, and S. Sorella, *J. Chem. Phys.* **121**, 7110 (2004).
- ⁷²W. G. Cochran, *Sampling Techniques*, 3rd ed. (John Wiley & Sons, 1977).
- ⁷³D. F. Gatz and L. Smith, *Atmos. Environ.* **29**, 1185 (1995).
- ⁷⁴H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.* **91**, 461 (1989).