

# Assessment of multireference approaches to explicitly correlated full configuration interaction quantum Monte Carlo

Cite as: J. Chem. Phys. **145**, 054117 (2016); <https://doi.org/10.1063/1.4959245>

Submitted: 23 May 2016 . Accepted: 10 July 2016 . Published Online: 04 August 2016

J. A. F. Kersten, George H. Booth, and Ali Alavi



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

[Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space](#)

The Journal of Chemical Physics **131**, 054106 (2009); <https://doi.org/10.1063/1.3193710>

[A deterministic alternative to the full configuration interaction quantum Monte Carlo method](#)

The Journal of Chemical Physics **145**, 044112 (2016); <https://doi.org/10.1063/1.4955109>

[Stochastic multi-reference perturbation theory with application to the linearized coupled cluster method](#)

The Journal of Chemical Physics **146**, 044107 (2017); <https://doi.org/10.1063/1.4974177>

Where in the **world** is AIP Publishing?

*Find out where we are exhibiting next*



# Assessment of multireference approaches to explicitly correlated full configuration interaction quantum Monte Carlo

J. A. F. Kersten,<sup>1,2,a)</sup> George H. Booth,<sup>3,b)</sup> and Ali Alavi<sup>1,2,c)</sup>

<sup>1</sup>University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, United Kingdom

<sup>2</sup>Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany

<sup>3</sup>Department of Physics, King's College London, Strand, London WC2R 2LS, United Kingdom

(Received 23 May 2016; accepted 10 July 2016; published online 4 August 2016)

The Full Configuration Interaction Quantum Monte Carlo (FCIQMC) method has proved able to provide near-exact solutions to the electronic Schrödinger equation within a finite orbital basis set, without relying on an expansion about a reference state. However, a drawback to the approach is that being based on an expansion of Slater determinants, the FCIQMC method suffers from a basis set incompleteness error that decays very slowly with the size of the employed single particle basis. The FCIQMC results obtained in a small basis set can be improved significantly with explicitly correlated techniques. Here, we present a study that assesses and compares two contrasting “universal” explicitly correlated approaches that fit into the FCIQMC framework: the  $[2]_{R12}$  method of Kong and Valeev [J. Chem. Phys. **135**, 214105 (2011)] and the explicitly correlated canonical transcorrelation approach of Yanai and Shiozaki [J. Chem. Phys. **136**, 084107 (2012)]. The former is an *a posteriori* internally contracted perturbative approach, while the latter transforms the Hamiltonian prior to the FCIQMC simulation. These comparisons are made across the 55 molecules of the G1 standard set. We found that both methods consistently reduce the basis set incompleteness, for accurate atomization energies in small basis sets, reducing the error from 28 mE<sub>h</sub> to 3–4 mE<sub>h</sub>. While many of the conclusions hold in general for any combination of multireference approaches with these methodologies, we also consider FCIQMC-specific advantages of each approach. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4959245>]

## I. INTRODUCTION

The vast majority of the properties of everyday molecular and solid state systems can be predicted from the solution of the non-relativistic, time-independent electronic Schrödinger equation. Static properties, which depend only on the ground state solution, cover important experimental quantities such as ionization potentials or atomization energies, required for heats of formation. However, the complexity of the fully coupled equations results in its exact solution being out of reach for all but the smallest systems to date. The Full Configuration Interaction Quantum Monte Carlo (FCIQMC) method<sup>1</sup> is a diagonalization-free, non-perturbative approach which projects a stochastic sampling of the wavefunction towards the exact ground state solution of the Schrödinger equation within a given orbital basis (OBS). This is achieved via a stochastic propagation of a population of walkers within the full space of Slater determinants. The dynamics are devised such that the long-time average over the discrete walker population on any determinant approaches that of the FCI coefficient. For several systems<sup>2,3</sup> it has been shown that the FCIQMC method is capable of providing the exact FCI energies within systematically improvable errorbars of  $O[10^{-4} - 10^{-5}]E_h$ , with only a fraction of the computational costs of an equivalent FCI calculation. Thus this method can

be applied to tackle difficult systems that require an accurate and non-perturbative description of the correlation. In the past, the FCIQMC method has been employed to calculate the energies of a diverse range of systems, including atoms and molecules,<sup>4–7</sup> the Hubbard model,<sup>8</sup> the uniform electron gas,<sup>9</sup> and solid state systems.<sup>10</sup> Since its first implementation, the efficiency and abilities of the FCIQMC method have been steadily improved. The initiator approximation,<sup>3,11,12</sup> real walkers, and semi-stochastic sampling<sup>13,14</sup> have decreased the computational effort of the method considerably, and new functionalities include the calculation of excited states,<sup>15–20</sup> the calculation of accurate two-particle reduced density matrices using a replica sampling technique,<sup>21,22</sup> and the use of these density matrices within a complete active space self-consistent field (CASSCF) approach.<sup>23,24</sup>

While an FCIQMC calculation is generally far less computationally expensive than a conventional FCI calculation, FCIQMC calculations are still limited to relatively small orbital basis sets, because the number of Slater determinants scales binomially with their size. This results in basis set incompleteness error (BSIE) in computed energies and properties, compounded by the fact that this error decreases slowly with respect to increasing basis set size.<sup>25–28</sup> The origin of this error is the result of the inability of Slater determinants to model the electron-electron cusps in the wavefunction.<sup>29–31</sup> As described by Kato's cusp condition,<sup>32</sup> the cusp regions are defined where the inter-electronic distance  $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$  goes to zero. Away from this point, the wavefunction depends

<sup>a)</sup>Electronic mail: jennifer.kersten@cantab.net

<sup>b)</sup>Electronic mail: george.booth@kcl.ac.uk

<sup>c)</sup>Electronic mail: a.alavi@fkf.mpg.de

linearly on  $r_{12}$ . Such behaviour is very difficult to obtain from a superposition of smooth Slater determinants which are built from one-electron functions that are centred at the nuclei. Thus very large basis sets are required to provide the flexibility for an accurate description of these cusp regions.

An alternative to extrapolating this basis set incompleteness through often expensive calculations in large basis sets is given by explicitly correlated methods,<sup>33–35</sup> which can reduce the BSIE efficiently with often little additional cost compared to the original calculation. The general aim of these approaches is to augment the ansatz for the wavefunction with explicitly correlated geminals which contain terms that are linear in the inter-electronic distance  $r_{12}$  at the cusp positions. Solutions for a large number of technical challenges from the inclusion of these geminals have been addressed in the literature over the last few decades.<sup>36–45</sup> These most notably include the avoidance of up to four-electron integrals, which have been addressed with efficient resolution of the identity techniques, and an optimized Slater-type geminal form, which provides a more efficient description of the wider correlation hole about each cusp position.<sup>46–51</sup> This work has led to modern “F12” methods which are reliable, efficient, and robust.

Whilst earlier efforts focused on the combination of F12 methodology with single-reference quantum chemical techniques, there has been increasing emphasis on adapting the technology for use with multi-reference methods, including multireference configuration interaction<sup>52,53</sup> and multireference perturbation theories,<sup>51,54</sup> within a consistent, internally contracted framework. This has been an important advance, since even though multireference methods are required in the presence of stronger, static correlations in low-spin open-shell or transition state systems, for example, these correlations are present *in addition* to the cusp-dominated dynamic correlation required for quantitative accuracy. These already challenging systems therefore also inherit all of the same slow basis set convergence problems of single reference systems. This is further exacerbated by the generally far higher cost with respect to basis set size for these systems, and therefore their combination with explicitly correlated techniques is a significant improvement. However, a drawback of many of the multireference approaches to date is that they are tailored and embedded within each parent multireference method, and therefore new equations have to be derived and code implemented, for each new multireference method.

This issue is rectified by the introduction of two “universal” explicitly correlated multireference techniques as the topic of the current study, which can be simply applied to almost all electronic structure methods. Their combination with the FCIQMC method provides a powerful approach to multireference quantum chemistry, and their overall accuracy, as well as suitability for use with FCIQMC will be assessed. These approaches are the  $[2]_{R12}$  method of Torheyden, Kong, and Valeev,<sup>55,56</sup> which is briefly described in Section II B, and the canonically transcorrelated approach of Yanai *et al.*,<sup>57,58</sup> described in Section II C. These provide contrasting approaches to the challenge of a universal multireference F12 method, with the former providing an internally contracted, perturbative coupling of the geminals to a given multireference wavefunction, while the latter uses an initial guess density in

order to transcorrelate the Hamiltonian, effectively removing the cusp features from its solution, prior to the multireference treatment. These approaches differ substantially in the way that the geminals are coupled to the multireference wavefunction, and in whether the multireference wavefunction can relax due to their presence, with the former adhering to a more traditional “diagonalize-then-perturb” approach, while the latter is constructed on a “perturb-then-diagonalize” philosophy. Whilst the infinite basis limit is exact for both approaches, away from this limit, they are not expected to behave the same. Previous applications of these approaches to FCIQMC have been shown to improve the results significantly via the reduction in BSIE, however these studies were limited to very few systems and did not allow for a direct comparison between the two different methods.<sup>59,60</sup> In this paper, the two explicitly correlated FCIQMC approaches are assessed in an in-depth study using the 55 molecules of the G1 standard set,<sup>61,62</sup> and by using this large sample size it is possible to draw more general conclusions about the success and efficiency of both methods.

## II. METHODS

### A. FCIQMC

In the Full Configuration Interaction Quantum Monte Carlo (FCIQMC) method, the wave function  $|\Psi\rangle$  is expanded in a basis set of all  $N$ -electron Slater determinants  $\{|D_I\rangle\}$ ,

$$|\Psi\rangle = \sum_I C_I |D_I\rangle. \quad (1)$$

The coefficients  $C_I$  are coarse-grained and stochastically sampled using a population of walkers, where each walker is defined by its sign, weight, and the Slater determinant it is associated with. The walkers then evolve according to a set of coupled differential equations derived from the imaginary time Schrödinger equation, such that the long-time average of the signed weight of walkers  $n_I$  on a Slater determinant  $D_I$  is proportional to  $C_I$ ,

$$n_I(\tau + \Delta\tau) = n_I(\tau) - \Delta\tau \sum_I (H_{JI} - E_S \delta_{IJ}) n_I(\tau), \quad (2)$$

where  $H_{JI}$  are the matrix elements of the Hamiltonian operator,  $E_S$  an energy offset, and  $\Delta\tau$  a small interval in imaginary time.

Following Equation (2), three consecutive steps are performed every time step  $\Delta\tau$ .

- *Spawning.* Each occupied determinant  $D_I$  spawns new walkers onto  $\chi_I$  randomly chosen connected determinants  $D_{J \neq I}$ , with

$$\chi_I = \begin{cases} \lceil |n_I| \rceil & \text{with probability } |n_I| - \lfloor |n_I| \rfloor \\ \lfloor |n_I| \rfloor & \text{otherwise,} \end{cases} \quad (3)$$

where  $\lfloor x \rfloor$  denotes the largest integer that is not greater than  $x$ , and  $\lceil x \rceil$  denotes the smallest integer that is not smaller than  $x$ . The sign of the newly spawned walkers on a determinant  $D_J$  is the same as parent determinant if  $H_{J,I} < 0$  and opposite otherwise, and their weight is

given by

$$p_s(J|I) = \frac{\delta\tau|H_{J,I}|}{p_{\text{gen}}(J|I)}, \quad (4)$$

where  $p_{\text{gen}}(J|I)$  denotes the probability of choosing the determinant  $D_J$ . More details on the specifics of this spawning step can be found in Ref. 22.

- *Death/cloning.* The weight of walkers  $n_I$  on each determinant  $D_I$  is reduced by a stochastically realized amount given by

$$p_d = \delta\tau (H_{I,I} - E_S) n_I. \quad (5)$$

- *Annihilation.* At the end of each time step, the list of newly spawned walkers is merged with the list of the old walkers (for algorithmic details of how to implement this efficiently, see Ref. 63). Walkers that are assigned to the same determinant but have an opposite sign annihilate each other, so that all walkers on each occupied determinant have the same sign at the end of each iteration.

A typical FCIQMC calculation starts with one walker on the Hartree–Fock (HF) determinant and a fixed energy offset  $E_S = E_{\text{HF}}$ . Due to the fact that the lowest eigenvalue of the Hamiltonian must be below  $E_{\text{HF}}$ , the population of walkers grows exponentially until a specified number of walkers is reached. Then  $E_S$  is varied smoothly, so that the walker population becomes approximately constant. The ground state energy can be determined from the long-time average of the energy offset  $E_S$  and from the long-time average of the projected energy

$$\langle E_p \rangle = E_{\text{HF}} + \frac{\langle \sum_{J \neq 0} H_{0,J} n_J \rangle}{\langle n_0 \rangle}, \quad (6)$$

where  $\langle \dots \rangle$  denotes an imaginary-time average, and  $n_0$  denotes the weight of walkers on the reference wavefunction. Often this is taken to be the largest weighted single Slater determinant, but can also be an arbitrary linear combination of Slater determinants.<sup>14</sup> Once the wavefunction has converged, and is sampling the desired solution, it is also possible to accumulate the two-body reduced density matrix, which is sampled at the same time as the spawning steps above. As this is a non-linear function of a random variable (the sampled wavefunction), serious systematic errors can result in this sampling. This issue is remedied via the introduction of a second “replica” sampling of the wavefunction, which samples randomly from the same distribution of the ground state wavefunction, but is uncorrelated to the first. This allows for an unbiased sampling of the reduced density matrices from which many properties of the wavefunction derive. Whilst the sampling of the density matrices is now unbiased, there are computational overheads with its computation. The sampling of the second replica approximately doubles the computational overhead, both in terms of memory and processor time, while additional (non-distributed) memory is required to store the full two-body density matrix (currently in non-sparse form). More details on the sampling of the density matrices, as well as more details on the FCIQMC algorithm can be found in Refs. 1–3, 22, and 63.

## B. The $[2]_{R12}$ approach

The spin-free  $[2]_{R12}$  approach developed by Valeev and co-workers can be employed *a posteriori* to reduce the basis set incompleteness error of an arbitrary multi-reference calculation.<sup>55,56,64</sup> The  $[2]_{R12}$  correction to the energy is evaluated in an internally contracted, second-order perturbative fashion via the Hylleraas functional. The first order wave function is expanded as explicitly correlated geminal replacements, constrained to be orthogonal, two-electron excitations to a multi-configurational reference wave function  $|0\rangle$ . In contrast to the MRMP2-F12 ansatz developed by Ten-no,<sup>51</sup> the explicitly correlated  $[2]_{R12}$  wave function also includes semi-internal excitations into geminal functions, and so the full first-order wavefunction can be written as

$$|\psi^{(1)}\rangle = \hat{\Omega}^{(1)}|0\rangle \quad (7)$$

$$= \frac{1}{2} t_{rs}^{pq} \left( r_{\alpha'\beta'}^{rs} \hat{E}_{pq}^{\alpha'\beta'} + 2r_{\alpha'x}^{rs} \hat{E}_{pq}^{\alpha'x} \right) |0\rangle - \frac{1}{2} t_{rs}^{pq} \left( 2r_{\alpha'k}^{rs} \left( \gamma^{(-1)} \right)_j^i \Gamma_{pq}^{jk} \hat{E}_i^{\alpha'j} \right) |0\rangle, \quad (8)$$

where  $r$  are the matrix elements of the correlation factor,  $\gamma$  and  $\Gamma$  are the spin-free one-body and two-body reduced density matrices (RDMs) of the reference wave function, and  $\hat{E}$  are the spin-free excitation operators. The notation of the orbitals belonging to the different parts of the orbital space is shown in Table I, and all equations are written using the Einstein summation convention, i.e., repeated indices are implicitly summed.

The geminal coefficients  $t_{rs}^{pq}$  are fixed according to the SP ansatz so that they satisfy the cusp conditions,<sup>48</sup>

$$t_{rs}^{pq} = \frac{3}{8} \delta_r^p \delta_s^q + \frac{1}{8} \delta_r^q \delta_s^p. \quad (9)$$

The second order Hylleraas function yields

$$\mathcal{H}^{(2)} = \langle \Psi_1 | [\hat{F}_N, \hat{\Omega}^{(1)}] + \hat{\Omega}^{(1)} \hat{F}_N | 0 \rangle + 2 \langle 0 | \hat{H}^{(1)} \hat{\Omega}^{(1)} | 0 \rangle, \quad (10)$$

where  $\hat{F}_N$  is the normal ordered spin-averaged Fock operator.

The  $\langle \Psi_1 | \hat{\Omega}^{(1)} \hat{F}_N | 0 \rangle$  term contains 4-body RDM terms, but its calculation can be avoided with the assumption that the generalized Brillouin condition is valid

$$\mathcal{H}^{(2)} \approx \langle \Psi_1 | [\hat{F}_N, \hat{\Omega}^{(1)}] | 0 \rangle + 2 \langle 0 | \hat{H}^{(1)} \hat{\Omega}^{(1)} | 0 \rangle. \quad (11)$$

The matrix elements are evaluated using the expanded Wick’s theorem, and the resulting expressions depend on the 1-body, 2-body, and 3-body reduced density matrices. The 3-body RDM terms are approximated with 1-body and 2-body

TABLE I. Notation of the orbital space.

Orbital space	Notation
Correlated orbitals	$p, q, r, s, t, u, v, w$
Occupied orbitals	$i, j, k$
Orbital basis sets (OBS)	$x, y, z$
Complementary auxiliary basis sets (CABS)	$\alpha', \beta'$
Complete virtual space	$\alpha, \beta$
Formally complete basis set (CBS)	$\kappa, \lambda$



terms by neglecting the irreducible 3-body cumulant in the generalized normal ordering of Kutzelnigg-Mukherjee.<sup>65–67</sup> Additionally, all terms that are quadratic in the 2-body RDM cumulants are discarded. Also, the screening approximation is employed and all terms in which a geminal matrix element and another geminal matrix element or a Coulomb matrix element that are connected over a 2-body cumulant vanish as well. The correlation factor,

$$f_{12} = -\frac{1}{\gamma} e^{-\gamma r_{12}}, \quad (12)$$

is fitted using 6 Gaussian functions to simplify the evaluation of the integrals. Further details on these approximations, along with the final working equations of the theory can be seen in Refs. 55 and 56.

In our work, the 2-body RDMs are sampled with the FCIQMC method along with the correlation energy using the stand-alone code NECI.<sup>68</sup> Whilst we have also implemented our own program for the  $[2]_{R12}$  corrections,<sup>59</sup> in this work we use the interface to the  $[2]_{R12}$  implementation within MPQC,<sup>69,70</sup> reading in the sampled density matrices and orbital information using an interface developed by Roskop *et al.*<sup>64</sup>

### C. Canonical transcorrelation theory

“Transcorrelated” methods, where the Hamiltonian operator is transformed by a Jastrow-style operator which compensates part of the correlated physics, has a long history in electronic structure, starting with Hirschfelder in 1963,<sup>71</sup> and extended by Boys and Handy.<sup>72–74</sup> These approaches used a similarity transformation of the Hamiltonian operator which rendered the resulting operator non-Hermitian, and numerically problematic. These ideas have been further developed by other authors to avoid many of the original shortcomings.<sup>46,47,75–79</sup> More recently, building on the work of White,<sup>80</sup> Chan, Yanai, and coworkers developed a related approach, but where the operator (trans)correlating the Hamiltonian is unitary. This yields a now Hermitian effective Hamiltonian, and is called Canonical Transformation (CT) theory.<sup>58,81–88</sup>

In the explicitly correlated version of canonical transformation theory developed by Yanai and Shiozaki,<sup>57</sup> the parameters for the unitary transformation operator,  $e^{\hat{A}}$  with  $-\hat{A} = \hat{A}^\dagger$ , are obtained from the projection of a set of strongly orthogonal explicitly correlated geminal functions. These are found as

$$\hat{A} = \frac{1}{2} \bar{R}_{ij}^{\alpha\beta} (\hat{E}_{ij}^{\alpha\beta} - \hat{E}_{\alpha\beta}^{ij}), \quad (13)$$

$$\bar{R}_{ij}^{\alpha\beta} = \frac{3}{8} \langle \alpha\beta | \hat{Q}_{12} f_{12} | ij \rangle + \frac{1}{8} \langle \alpha\beta | \hat{Q}_{12} f_{12} | ji \rangle, \quad (14)$$

$$\hat{Q}_{12} = (1 - \hat{O}_1)(1 - \hat{O}_2) - \hat{V}_1 \hat{V}_2, \quad (15)$$

where the one-electron operators  $\hat{O}$  and  $\hat{V}$  project onto the occupied orbitals and the virtual orbitals of the OBS, respectively, the SP ansatz has been used to fix the geminal coefficients, and the same Slater-type geminal form is used as in Eq. (12). The effective Hamiltonian is derived as

$$\hat{H}_{TC} = e^{\hat{A}^\dagger} \hat{H} e^{\hat{A}} \quad (16)$$

$$= \hat{H} + [\hat{H}, \hat{A}] + \frac{1}{2!} [[\hat{H}, \hat{A}], \hat{A}] + \dots, \quad (17)$$

where  $\hat{H}_{TC}$  is Hermitian, and the transcorrelated eigenvalue problem can be solved with the traditional post-HF methods that are built on the variational principle.

In order to simplify the calculation, several approximations are made for the transcorrelated Hamiltonian. Firstly, the Baker-Campbell-Hausdorff expansion is truncated after the second order. Secondly, all commutators are approximated with one- and two-body operators, denoted with  $[\dots]_{1,2}$ . Similar to the decomposition and approximation of the 3-particle density matrix using the Kutzelnigg-Mukherjee formalism, the higher order excitation operators are replaced with an approximate decomposition into one- and two-body terms. In addition, in the second order term, the Hamiltonian is replaced by the Fock operator  $\hat{F}$ ,

$$\hat{H}_{TC} \approx \hat{H} + [\hat{H}, \hat{A}]_{1,2} + \frac{1}{2!} [[\hat{F}, \hat{A}], \hat{A}]_{1,2}. \quad (18)$$

The resulting terms for the matrix elements of the transcorrelated Hamiltonian are similar to the intermediates emerging in standard F12 calculations and the many-electron integrals are solved in a similar manner using standard approximation C,<sup>89</sup> and RI insertions, resulting in an effective Hamiltonian which is only two-body, and can therefore be treated by standard electronic structure methods.

The electron density which is used in order to define the initial geminal functions must be supplied prior to the calculation of the effective Hamiltonian, technically resulting in a loss of its state-universal characteristics. In this work, we use a trial density obtained from a prior CASSCF calculation, in order to allow the geminals to be constructed in the presence of any static correlation. The matrix elements of the transcorrelated Hamiltonian are then computed with the stand-alone code ORZ, using the F12 integral engine of Shiozaki,<sup>90–92</sup> constructing the geminal functions within the same active space that was chosen for the CASSCF calculation. Since the resulting transcorrelated Hamiltonian is Hermitian and of two-particle form, the FCIQMC method is then employed to find the lowest energy eigenvalue.

## III. RESULTS

In order to assess the quality of the two explicitly correlated FCIQMC approaches, we considered the calculation of total energies and atomization energies of the 55 molecules across the G1 standard set.<sup>61,62</sup> FCIQMC used in conjunction with the  $[2]_{R12}$  method is denoted FCIQMC-R12 and the canonical transcorrelation approach with CT-FCIQMC. The G1 molecules are relatively small, single-reference dominated systems that have been extensively studied in the past.<sup>93–98</sup> All geometries were fully optimized with GAMESS at the MP2/aug-cc-pVQZ level. All other calculations were performed with frozen cores to reduce the computational effort.

The FCIQMC calculations were performed with the initiator approximation,<sup>12</sup> the semi-stochastic method,<sup>13,14</sup> and the replica sampling where the RDM is required.<sup>21,22</sup> The size of the deterministic space for the semi-stochastic adaptation was chosen to be one tenth of the size of the initiator

space unless this was prohibited by memory limitations ( $O[10^4 - 10^5]$  determinants). For the closed shell systems, time-reversal symmetry was also employed.<sup>3</sup> We obtained the correlation energies and statistical errors from Flyvbjerg-Petersen blocking analyses of the projected energies.<sup>99</sup> For all cases, the statistical errors were found to be very small ( $O[10^{-5}]E_h$ ), and so statistical error bars are generally not visible in the following figures. The leading error is hence the initiator error, which can be systematically improved by increasing the total number of walkers. For the molecules of the G1 standard set, the total number of walkers was chosen individually, depending on the convergence of the system, between  $50 \times 10^6$  and  $1 \times 10^9$  walkers. While the chosen number of walkers was sufficient to ensure that the initiator error in the energies of the smaller systems is negligible, it cannot be guaranteed that the energies of the larger molecules are entirely free from initiator error. Quantifying the remaining initiator error is difficult, but we estimate it to be  $O[10^{-4}]E_h$  for the total energy of the largest systems. This is certainly small enough not to affect any resulting conclusions or qualitative trends in the results. For the explicitly correlated corrections, we set the parameter  $\gamma$  to  $1 a_0^{-1}$ , and used the aug-cc-pVDZ-OPTRI basis sets for the construction of the CABS where possible.

Along with incompleteness in the description of two-electron part of the wavefunction, there is also a basis set incompleteness in the one-electron description of the wavefunction,<sup>35,59,100</sup> which is excluded from the construction of these universal F12 corrections. While the presence of strong correlations can change the magnitude of this one-electron incompleteness, it is generally much smaller than the two-electron incompleteness, and so we assume here that it is well represented by the incompleteness in the uncorrelated Hartree–Fock energy. To estimate the complete basis set (CBS) limit of this one-electron energy, we extrapolated the HF energies to the CBS limit with an exponential dependence,<sup>101</sup> as

$$E_{\text{CBS}}^{\text{HF}} \approx \frac{E_X^{\text{HF}} - bE_{X-1}^{\text{HF}}}{1 - b}, \quad (19)$$

with

$$b = \frac{E_X^{\text{HF}} - E_{X-1}^{\text{HF}}}{E_{X-1}^{\text{HF}} - E_{X-2}^{\text{HF}}}, \quad (20)$$

where  $X$  is the cardinal number of the basis set and calculations in aug-cc-pVTZ, aug-cc-pVQZ, and aug-cc-pV5Z basis sets were used. We note that it is possible for other methods to be used to estimate the one-electron basis set incompleteness from correlated wavefunctions.<sup>102</sup>

## A. Total energies

The molecules of the G1 standard set have a generally single-reference character so that CCSD(T) calculations are capable of achieving high accuracy results. It is likely then that these results will be a fairer comparison than experimentally derived results for the total energy, due to uncertainties in the experimental procedure for this quantity. Hence, the assumed exact energies in the CBS limit that we used as a benchmark were obtained from CCSD(T)-F12b calculations in aug-cc-pV5Z basis sets performed with MOLPRO,<sup>100,103,104</sup> which we then further correct for static correlation effects. The static correlation corrections for the different molecules were approximated with the individual error of a CCSD(T) calculation in an aug-cc-pVDZ basis set compared to the near-exact FCIQMC result in the same basis set. Since the majority of the static components of correlation can be captured in a small orbital space (hence the success of the CASSCF-derived approaches to strongly correlated systems), it is reasonable to assume that this correction will account for most of the deficiencies in the CCSD(T) results. As expected, the static corrections are small for those single-reference systems: The absolute mean error is  $0.1 \text{ mE}_h$  per correlated electron, and the maximum absolute total error of  $3.0 \text{ mE}_h$  ( $0.4 \text{ mE}_h$  per correlated electron) was found for the  $\text{Si}_2$  molecule.

In Fig. 1 we compare the FCIQMC results in aug-cc-pVDZ basis sets with the results from FCIQMC-R12 and CT-FCIQMC calculations with added one-electron BSIE corrections for all 55 molecules of the G1 standard set. It

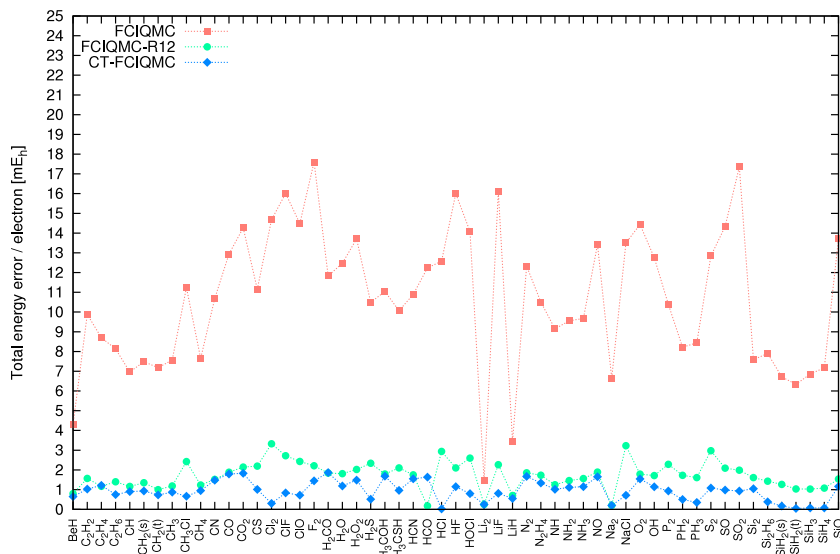


FIG. 1. Absolute errors of the total energies per correlated electron for the G1 set of molecules, calculated with the FCIQMC method, and the FCIQMC-R12 and CT-FCIQMC methods (with one-electron BSIE corrections) in aug-cc-pVDZ basis sets.

can be seen that both explicitly correlated methods reduce the absolute error per correlated electron significantly for all molecules. The effect of the correction is stronger for the systems that show a larger BSIE, such as  $\text{SO}_2$  and  $\text{F}_2$ . Furthermore, for nearly all molecules, the CT-FCIQMC method achieves slightly lower errors than the FCIQMC-R12 method. On average, the CT method and the  $[2]_{R12}$  method reduce the absolute error in the total energy per correlated electron in an aug-cc-pVDZ basis from 10.8  $\text{mE}_h$  to 1.0  $\text{mE}_h$  and 1.8  $\text{mE}_h$ , respectively.

## B. The atomization energies

We also compared the performance of the explicitly correlated corrections for the atomization energies. As benchmark results, we use experimental data<sup>105–120</sup> corrected for zero-point vibrations, spin-orbit coupling, core-valence effects, and scalar relativistic effects.<sup>121</sup> In Fig. 2 the absolute error is shown for the atomization energy of each molecule. Again, we can see that the corrections for the basis set incompleteness error improve the FCIQMC results drastically. For this case of relative energies, the difference between the CT and the  $[2]_{R12}$  method is small.

To better analyse and compare the distribution of errors between the different methods, Fig. 3 shows the density of the errors in the atomization energies achieved with different methods fitted to a Weibull distribution.<sup>122,123</sup> We can clearly see that the median of the error distribution across the set is reduced from 28.2  $\text{mE}_h$  to about 2.9  $\text{mE}_h$  with the CT correction and to 4.1  $\text{mE}_h$  with the  $[2]_{R12}$  correction. Thus, the FCIQMC-R12 and the FCIQMC-CT methods are capable of producing relatively accurate atomization energies in aug-cc-pVDZ basis sets (though not quite chemical accuracy), with far greater reliability than without the explicit correlation, indicated by the smaller spread and median in the results. It is notable that the spread and the median in these errors are slightly larger for the FCIQMC-R12 method, compared to the CT method, suggesting a marginally less reliable reduction in the error with the approach.

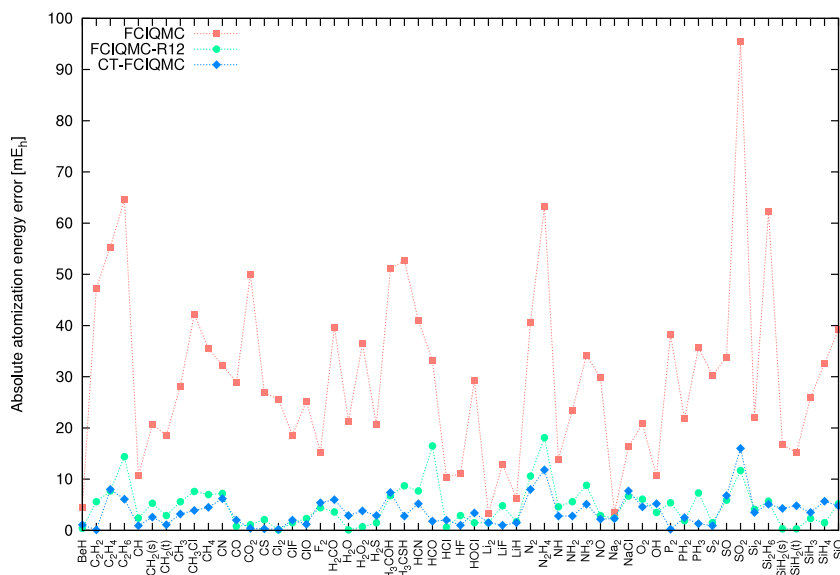


FIG. 2. Absolute errors of the atomization energies for the G1 set of molecules, calculated with the FCIQMC method, and the FCIQMC-R12 and CT-FCIQMC methods (with one-electron BSIE corrections) in aug-cc-pVDZ basis sets.

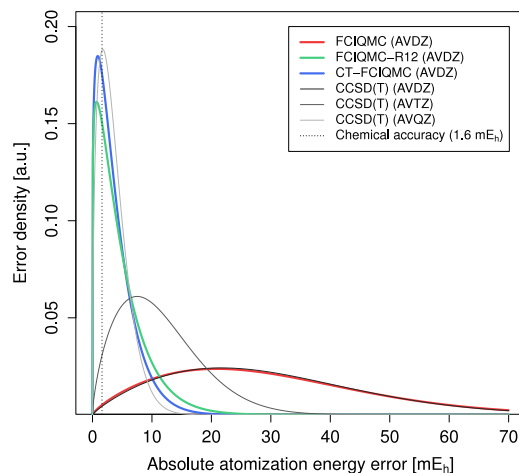


FIG. 3. Distribution of the absolute errors in the atomization energies of the G1 set, fitted to Weibull distributions.

Fig. 3 also shows the atomization error distributions obtained with CCSD(T) calculations without explicit correlation in increasingly large basis sets. The CCSD(T) results are mostly very accurate within the given basis set for the molecules of the G1 standard set, as shown by the similarity of the CCSD(T) and FCIQMC distributions for the aug-cc-pVDZ basis. However, as has been noted elsewhere, the explicitly correlated calculations in aug-cc-pVDZ basis sets achieve a similar accuracy as CCSD(T) calculations in aug-cc-pVQZ basis sets, already without including the additional static correlation component of the electronic structure which is obtained via the FCIQMC. Thus it can be concluded that the R12 and the CT method both effectively decrease the BSIE to a level that is normally achieved with basis sets that are two cardinal numbers higher.

## C. The initiator error

The initiator approximation reduces the computational effort of an FCIQMC calculation significantly, but introduces

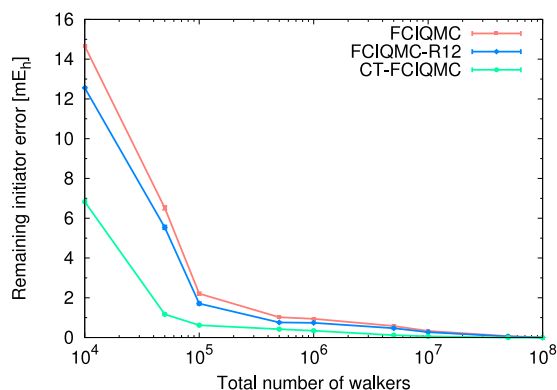


FIG. 4. Decay of the initiator error with respect to the total number of walkers for the FCIQMC, the FCIQMC-R12, and the CT-FCIQMC approaches for the  $N_2$  molecule in an aug-cc-pVDZ basis.

an initiator error into the sampled energies. This error can be reduced systematically by increasing the total number of walkers, and the nature of this convergence is illustrated for the FCIQMC, FCIQMC-R12, and CT-FCIQMC methods for the  $N_2$  molecule in Fig. 4. It can be seen that the CT-FCIQMC energy converges noticeably faster than the FCIQMC energy with respect to the total number of walkers. To reach convergence within 1 mE<sub>h</sub>, the CT-FCIQMC calculation requires less than  $10^5$  walkers while the conventional FCIQMC calculation needs more than  $5 \times 10^5$  walkers to achieve the same level of convergence. This indicates that removing the short-ranged part of the Coulomb hole from the Hamiltonian improves the convergence of the initiator error, and fewer walkers and thus less computational effort is required to determine the ground state energy of the transcorrelated Hamiltonian accurately. This is despite having the same number of degrees of freedom and overall Slater determinants as the other calculations. The  $[2]_{R12}$  correction requires the reduced density matrix from the FCIQMC calculation and therefore suffers from the initiator error as well. However, the error in the  $[2]_{R12}$  correction is small compared to the basis set correlation energy and decays quickly with respect to the total number of walkers. However, it also has the opposite sign from the error in the FCIQMC energy, and thus slightly reduces the overall initiator error at low walker numbers due to favourable error cancellation.

#### IV. CONCLUSIONS

The canonical transcorrelation approach of Yanai and Shiozaki and the  $[2]_{R12}$  approach of Valeev *et al.* constitute two contrasting philosophies in the attempt to create a universal, explicitly correlated approach to basis set incompleteness in quantum chemistry. In the former, the dynamic correlation from the cusps is included first, to produce an effective, two-body Hamiltonian operator, to which the multireference strong correlation methods can be applied. This has the advantage that the multireference wavefunction capturing the static correlation in the system can relax in the presence of the geminal amplitudes and dynamic correlation contributions. Furthermore, we showed that the transcorrelation reduces the

computational effort of the FCIQMC calculation as it improves the convergence of the initiator error with respect to the total number of walkers. The CT method is also in principle a state-universal approach, in that the specifics of the high-energy geminal contributions to the effective Hamiltonian should be relatively insensitive to the state or specifics of the choice of geminal. However, one source of ambiguity in the construction of  $\hat{H}_{TC}$  comes from the choice of the active space that is required for the CASSCF calculation of the initial “trial” density, and for the transcorrelation itself.

Contrasting this, the  $[2]_{R12}$  perturbatively couples the geminals directly to the multireference wavefunction, which provides geminal relaxation in a state-specific fashion and without any initial trial wavefunction, but conversely does not allow for relaxation of the CI amplitudes in the presence of dynamic correlation component of the electronic structure. However, these effects are likely to be minor details in the largely single-reference systems studied here, and it can be seen that in both methods, the FCIQMC-R12 and CT-FCIQMC methods both achieve accurate results even in the small basis sets, with the CT approach slightly outperforming the  $[2]_{R12}$  method. Computational requirements for the integration of the  $[2]_{R12}$  methodology with FCIQMC are larger, due to the necessity of sampling the two-body density matrix. The sampling at least doubles the memory and processing cost of the FCIQMC calculation due to the requirement of two replicas.

However, aside from the cost of the density matrix sampling, the computational effort of both methods is very cheap compared to the costs of the FCIQMC calculations, scaling as  $O[N^6]$  with the system size. The different calculations for the molecules of the G1 standard set took between a few minutes and about two days on a single core for both of the explicitly correlated treatments. Conclusively, with little additional computational effort, explicitly correlated FCIQMC calculations in small basis sets can achieve the same accuracy that is normally only achieved with considerably more expensive FCIQMC calculations in larger basis sets. This broadens the applicability of the FCIQMC method considerably and allows accurate calculations of larger systems, where in the future, we will consider problems with stronger correlation, and applications to excited states and solid state systems.<sup>124–126</sup>

#### ACKNOWLEDGMENTS

We would like to sincerely thank Takeshi Yanai for kindly allowing us to use his code ORZ, and are also grateful to Luke Roskop for his help with the MPQC-GAMESS interface. G.H.B. gratefully acknowledges funding from the Royal Society. The work has been funded by the EPSRC under Grant No. EP/J003867/1.

<sup>1</sup>G. H. Booth, A. J. W. Thom, and A. Alavi, *J. Chem. Phys.* **131**, 054106 (2009).

<sup>2</sup>G. H. Booth and A. Alavi, *J. Chem. Phys.* **132**, 174104 (2010).

<sup>3</sup>G. H. Booth, D. Cleland, A. J. W. Thom, and A. Alavi, *J. Chem. Phys.* **135**, 084104 (2011).

<sup>4</sup>R. E. Thomas, C. Overy, G. H. Booth, and A. Alavi, *J. Chem. Theor. Comput.* **10**, 1915 (2014).



- <sup>5</sup>C. Daday, S. Smart, G. H. Booth, A. Alavi, and C. Filippi, *J. Chem. Theor. Comput.* **8**, 4441 (2012).
- <sup>6</sup>D. Cleland, G. H. Booth, C. Overy, and A. Alavi, *J. Chem. Theor. Comput.* **8**, 4138 (2012).
- <sup>7</sup>R. E. Thomas, G. H. Booth, and A. Alavi, *Phys. Rev. Lett.* **114**, 033001 (2015).
- <sup>8</sup>L. R. Schwarz, G. H. Booth, and A. Alavi, *Phys. Rev. B* **91**, 045139 (2015).
- <sup>9</sup>J. J. Shepherd, G. H. Booth, A. Grüneis, and A. Alavi, *Phys. Rev. B* **85**, 081103(R) (2012).
- <sup>10</sup>G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, *Nature* **493**, 365 (2013).
- <sup>11</sup>D. Cleland, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **132**, 041103 (2010).
- <sup>12</sup>D. M. Cleland, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **134**, 024112 (2011).
- <sup>13</sup>F. R. Petruzielo, A. A. Holmes, H. J. Changlani, M. P. Nightingale, and C. J. Umrigar, *Phys. Rev. Lett.* **109**, 230201 (2012).
- <sup>14</sup>N. S. Blunt, S. D. Smart, J. A. F. Kersten, J. S. Spencer, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **142**, 184107 (2015).
- <sup>15</sup>A. Humeniuk and R. Mitrić, *J. Chem. Phys.* **141**, 194104 (2014).
- <sup>16</sup>N. S. Blunt, A. Alavi, and G. H. Booth, *Phys. Rev. Lett.* **115**, 050603 (2015).
- <sup>17</sup>G. H. Booth and G.-L. Chan, *J. Chem. Phys.* **137**, 191102 (2012).
- <sup>18</sup>Y. Ohtsuka and S. Nagase, *Theor. Chem. Acc.* **130**, 501 (2011).
- <sup>19</sup>S. Ten-no, *J. Chem. Phys.* **138**, 164126 (2013).
- <sup>20</sup>N. Blunt, S. D. Smart, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **143**, 134117 (2015).
- <sup>21</sup>N. S. Blunt, T. W. Rogers, J. S. Spencer, and W. M. C. Foulkes, *Phys. Rev. B* **89**, 245124 (2014).
- <sup>22</sup>C. Overy, G. H. Booth, N. S. Blunt, J. J. Shepherd, D. Cleland, and A. Alavi, *J. Chem. Phys.* **141**, 244117 (2014).
- <sup>23</sup>R. E. Thomas, Q. Sun, A. Alavi, and G. H. Booth, *J. Chem. Theor. Comput.* **11**, 5316 (2015).
- <sup>24</sup>G. L. Manni, S. D. Smart, and A. Alavi, *J. Chem. Theor. Comput.* **12**, 1245 (2016).
- <sup>25</sup>C. Schwartz, *Phys. Rev.* **126**, 1015 (1962).
- <sup>26</sup>R. N. Hill, *J. Chem. Phys.* **83**, 1173 (1985).
- <sup>27</sup>W. Kutzelnigg and J. Morgan III, *J. Chem. Phys.* **96**, 4484 (1992).
- <sup>28</sup>J. J. Shepherd, A. Grüneis, G. H. Booth, G. Kresse, and A. Alavi, *Phys. Rev. B* **86**, 035111 (2012).
- <sup>29</sup>J. C. Slater, *Phys. Rev.* **31**, 333 (1928).
- <sup>30</sup>E. A. Hylleraas, *Z. Phys.* **54**, 347 (1929).
- <sup>31</sup>R. T Pack and W. B. Brown, *J. Chem. Phys.* **45**, 556 (1966).
- <sup>32</sup>T. Kato, *Commun. Pure Appl. Math.* **10**, 151 (1957).
- <sup>33</sup>W. Klopper, F. R. Manby, S. Ten-No, and E. F. Valeev, *Int. Rev. Phys. Chem.* **25**, 427 (2006).
- <sup>34</sup>L. Kong, F. A. Bischoff, and E. F. Valeev, *Chem. Rev.* **112**, 75 (2012).
- <sup>35</sup>C. Hättig, W. Klopper, A. Köhn, and D. P. Tew, *Chem. Rev.* **112**, 4 (2012).
- <sup>36</sup>W. Kutzelnigg, *Theor. Chim. Acta* **68**, 445 (1985).
- <sup>37</sup>V. Termath, W. Klopper, and W. Kutzelnigg, *J. Chem. Phys.* **94**, 2002 (1991).
- <sup>38</sup>W. Kutzelnigg and W. Klopper, *J. Chem. Phys.* **94**, 1985 (1991).
- <sup>39</sup>J. Noga, W. Kutzelnigg, and W. Klopper, *Chem. Phys. Lett.* **199**, 497 (1992).
- <sup>40</sup>J. Noga and W. Kutzelnigg, *J. Chem. Phys.* **101**, 7738 (1994).
- <sup>41</sup>W. Klopper and C. C. M. Samson, *J. Chem. Phys.* **116**, 6397 (2002).
- <sup>42</sup>E. F. Valeev, *Chem. Phys. Lett.* **395**, 190 (2004).
- <sup>43</sup>A. J. May and F. R. Manby, *J. Chem. Phys.* **121**, 4479 (2004).
- <sup>44</sup>F. R. Manby, *J. Chem. Phys.* **119**, 4607 (2003).
- <sup>45</sup>S. Ten-no, *J. Chem. Phys.* **126**, 014108 (2007).
- <sup>46</sup>S. Ten-no, *Chem. Phys. Lett.* **330**, 169 (2000).
- <sup>47</sup>S. Ten-no, *Chem. Phys. Lett.* **330**, 175 (2000).
- <sup>48</sup>S. Ten-no, *Chem. Phys. Lett.* **398**, 56 (2004).
- <sup>49</sup>D. P. Tew and W. Klopper, *J. Chem. Phys.* **123**, 074101 (2005).
- <sup>50</sup>A. J. May, E. Valeev, R. Polly, and F. R. Manby, *Phys. Chem. Chem. Phys.* **7**, 2710 (2005).
- <sup>51</sup>S. Ten-no, *Chem. Phys. Lett.* **447**, 175 (2007).
- <sup>52</sup>R. J. Gdanitz, *Chem. Phys. Lett.* **312**, 578 (1999).
- <sup>53</sup>T. Shiozaki, G. Knizia, and H.-J. Werner, *J. Chem. Phys.* **134**, 034113 (2011).
- <sup>54</sup>T. Shiozaki and H.-J. Werner, *J. Chem. Phys.* **133**, 141103 (2010).
- <sup>55</sup>M. Torheyden and E. F. Valeev, *J. Chem. Phys.* **131**, 171103 (2009).
- <sup>56</sup>L. Kong and E. F. Valeev, *J. Chem. Phys.* **135**, 214105 (2011).
- <sup>57</sup>T. Yanai and T. Shiozaki, *J. Chem. Phys.* **136**, 084107 (2012).
- <sup>58</sup>T. Yanai and G. K.-L. Chan, *J. Chem. Phys.* **127**, 104107 (2007).
- <sup>59</sup>G. H. Booth, D. Cleland, A. Alavi, and D. P. Tew, *J. Chem. Phys.* **137**, 164112 (2012).
- <sup>60</sup>S. Sharma, T. Yanai, G. H. Booth, C. J. Umrigar, and G. K.-L. Chan, *J. Chem. Phys.* **140**, 104112 (2014).
- <sup>61</sup>J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, *J. Chem. Phys.* **90**, 5622 (1989).
- <sup>62</sup>L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.* **93**, 2537 (1990).
- <sup>63</sup>G. H. Booth, S. D. Smart, and A. Alavi, *Mol. Phys.* **112**, 1855 (2014).
- <sup>64</sup>L. B. Roskop, L. Kong, E. F. Valeev, M. S. Gordon, and T. L. Windus, *J. Chem. Theor. Comput.* **10**, 90 (2014).
- <sup>65</sup>W. Kutzelnigg and D. Mukherjee, *J. Chem. Phys.* **107**, 432 (1997).
- <sup>66</sup>W. Kutzelnigg, K. R. Shamasundar, and D. Mukherjee, *Mol. Phys.* **108**, 433 (2010).
- <sup>67</sup>W. Kutzelnigg, *J. Chem. Phys.* **125**, 171101 (2006).
- <sup>68</sup>The FCIQMC code can be obtained from [https://github.com/ghb24/NECI\\_STABLE.git](https://github.com/ghb24/NECI_STABLE.git).
- <sup>69</sup>C. L. Janssen, I. B. Nielsen, M. L. Leininger, E. F. Valeev, J. P. Kenny, and E. T. Seidl, The massively parallel quantum chemistry program (MPQC), version 2.4.0, 2006.
- <sup>70</sup>C. Janssen, E. Seidl, and M. Colvin, in *ACS Symposium Series, Parallel Computing in Computational Chemistry* (American Chemical Society, 1995), Vol. 592.
- <sup>71</sup>J. O. Hirschfelder, *J. Chem. Phys.* **39**, 3145 (1963).
- <sup>72</sup>S. F. Boys and N. C. Handy, *Proc. R. Soc. London, Ser. A* **309**, 209 (1969).
- <sup>73</sup>S. F. Boys and N. C. Handy, *Proc. R. Soc. London, Ser. A* **310**, 43 (1969).
- <sup>74</sup>S. F. Boys and N. C. Handy, *Proc. R. Soc. London, Ser. A* **310**, 63 (1969).
- <sup>75</sup>N. Umezawa and S. Tsuneyuki, *J. Chem. Phys.* **119**, 10015 (2003).
- <sup>76</sup>N. Umezawa and S. Tsuneyuki, *Phys. Rev. B* **69** (2004).
- <sup>77</sup>H. Luo, *J. Chem. Phys.* **133**, 154109 (2010).
- <sup>78</sup>H. Luo, *J. Chem. Phys.* **135**, 024109 (2011).
- <sup>79</sup>H. Luo, *J. Chem. Phys.* **136**, 224111 (2012).
- <sup>80</sup>S. R. White, *J. Chem. Phys.* **117**, 7472 (2002).
- <sup>81</sup>E. Neuscammann, T. Yanai, and G.-L. Chan, *Int. Rev. Phys. Chem.* **29**, 231 (2010).
- <sup>82</sup>T. Yanai and G.-L. Chan, *J. Chem. Phys.* **124**, 194106 (2006).
- <sup>83</sup>G.-L. Chan and T. Yanai, *Adv. Chem. Phys.* **134**, 343 (2007).
- <sup>84</sup>E. Neuscammann, T. Yanai, and G.-L. Chan, *J. Chem. Phys.* **130**, 124102 (2009).
- <sup>85</sup>E. Neuscammann, T. Yanai, and G.-L. Chan, *J. Chem. Phys.* **132**, 024106 (2010).
- <sup>86</sup>T. Yanai, Y. Kurashige, E. Neuscammann, and G.-L. Chan, *J. Chem. Phys.* **132**, 024105 (2010).
- <sup>87</sup>T. Yanai, Y. Kurashige, E. Neuscammann, and G.-L. Chan, *Phys. Chem. Chem. Phys.* **14**, 7809 (2012).
- <sup>88</sup>T. Watson and G.-L. Chan, *J. Chem. Theor. Comput.* **12**, 512 (2016).
- <sup>89</sup>S. Kedzuch, M. Milko, and J. Noga, *Int. J. Quantum Chem.* **105**, 929 (2005).
- <sup>90</sup>T. Shiozaki, *Chem. Phys. Lett.* **479**, 160 (2009).
- <sup>91</sup>T. Shiozaki, *J. Unsolved Quest.* **1**, 1 (2011), <http://junq.info/?p=348>; accessed on 30 December 2011.
- <sup>92</sup>T. Shiozaki, POLYR12, the program for explicitly correlated electron-correlation calculations of molecules and solids, can be obtained from <https://github.com/nubakery/libslater>, 2014.
- <sup>93</sup>D. Feller, K. A. Peterson, and D. A. Dixon, *J. Chem. Phys.* **129**, 204105 (2008).
- <sup>94</sup>J. C. Grossman, *J. Chem. Phys.* **117**, 1434 (2002).
- <sup>95</sup>A. D. Becke, *J. Chem. Phys.* **96**, 2155 (1992).
- <sup>96</sup>M. Ernzerhof and G. E. Scuseria, *J. Chem. Phys.* **110**, 5029 (1999).
- <sup>97</sup>N. Nemec, M. D. Towler, and R. J. Needs, *J. Chem. Phys.* **132**, 034111 (2010).
- <sup>98</sup>M. A. Morales, J. McMinis, B. K. Clark, J. Kim, and G. E. Scuseria, *J. Chem. Theor. Comput.* **8**, 2181 (2012).
- <sup>99</sup>H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.* **91**, 461 (1989).
- <sup>100</sup>T. B. Adler, G. Knizia, and H.-J. Werner, *J. Chem. Phys.* **127**, 221106 (2007).
- <sup>101</sup>A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, *Chem. Phys. Lett.* **302**, 437 (1999).
- <sup>102</sup>L. Kong and E. F. Valeev, *J. Chem. Phys.* **133**, 174126 (2010).
- <sup>103</sup>H. Werner, P. Knowles, G. Knizia, F. Manby, M. Schütz *et al.*, MOLPRO, version 2012.1, a package of *ab initio* programs, 2012, see <http://www.molpro.net>.
- <sup>104</sup>G. Knizia, T. B. Adler, and H.-J. Werner, *J. Chem. Phys.* **130**, 054104 (2009).
- <sup>105</sup>M. Chase, Jr. and N.-J. T. Tables, *J. Phys. Chem. Ref. Data* **1** (1998).
- <sup>106</sup>K. Huber and G. Herzberg, *Constants of Diatomic Molecules, Molecular Spectra and Molecular Structure* Vol. 4 (Van Nostrand Reinhold, New York, 1979).
- <sup>107</sup>K. R. Way and W. C. Stwalley, *J. Chem. Phys.* **59**, 5298 (1973).
- <sup>108</sup>Y. Huang, S. A. Barts, and J. B. Halpern, *J. Phys. Chem.* **96**, 425 (1992).

- <sup>109</sup>R. Lengel and R. Zare, *J. Am. Chem. Soc.* **100**, 7495 (1978).
- <sup>110</sup>A. McKellar, P. Bunker, T. J. Sears, K. Evenson, R. J. Saykally, and S. Langhoff, *J. Chem. Phys.* **79**, 5251 (1983).
- <sup>111</sup>D. Baulch, R. Cox, P. Crutzen, R. Hampson, Jr., J. Kerr, J. Troe, R. Watson *et al.*, *J. Phys. Chem. Ref. Data* **11**, 327 (1982).
- <sup>112</sup>S. Gibson, J. Greene, and J. Berkowitz, *J. Chem. Phys.* **83**, 4319 (1985).
- <sup>113</sup>D. D. Wagman, W. H. Evans, V. B. Parker, R. H. Schumm, I. Halow, S. M. Bailey, K. L. Churney, and R. L. Nuttall, *J. Phys. Chem. Ref. Data, Suppl.* **2**, 11 (1982).
- <sup>114</sup>S. Lias, J. Bartmess, J. Liebman, J. Holmes, R. Levin, and W. G. Mallard, Ion Energetics Data, NIST Chemistry Web-Book, NIST Standard Reference Database Number 69, edited by P. J. Linstrom and W. G. Mallard, National Institute of Standards and Technology, Gaithersburg MD 20899, 1988.
- <sup>115</sup>V. P. Glushko, L. V. Gurvich, G. A. Bergman, I. V. Veitz, V. A. Medvedev, G. A. Khachkuruzov, and V. S. Jungman, *Thermodynamic Properties of Pure Substances* (Nauka, Moscow, USSR, 1978).
- <sup>116</sup>J. Berkowitz, J. P. Greene, H. Cho, and B. Rui, *J. Chem. Phys.* **86**, 1235 (1987).
- <sup>117</sup>J. E. Butler, K. Kawaguchi, and E. Hirota, *J. Mol. Spectrosc.* **101**, 161 (1983).
- <sup>118</sup>J. Berkowitz, L. Curtiss, S. Gibson, J. Greene, G. Hillhouse, and J. Pople, *J. Chem. Phys.* **84**, 375 (1986).
- <sup>119</sup>N. S. Dattani, *J. Mol. Spectrosc.* **311**, 76 (2015).
- <sup>120</sup>R. J. Le Roy, N. S. Dattani, J. A. Coxon, A. J. Ross, P. Crozet, and C. Linton, *J. Chem. Phys.* **131**, 204309 (2009).
- <sup>121</sup>D. Feller and K. A. Peterson, *J. Chem. Phys.* **110**, 8384 (1999).
- <sup>122</sup>M. Fréchet, *Annales de la Societe Polonaise de Mathematique* **6**, 93 (1927).
- <sup>123</sup>W. Weibull, *ASME J. Appl. Mech.* **18**, 293 (1951).
- <sup>124</sup>A. Gruneis, J. J. Shepherd, A. Alavi, P. T. Tew, and G. H. Booth, *J. Chem. Phys.* **139**, 084112 (2013).
- <sup>125</sup>A. Gruneis, *Phys. Rev. Lett.* **115**, 066402 (2015).
- <sup>126</sup>D. Usvyat, *J. Chem. Phys.* **139**, 194101 (2013).