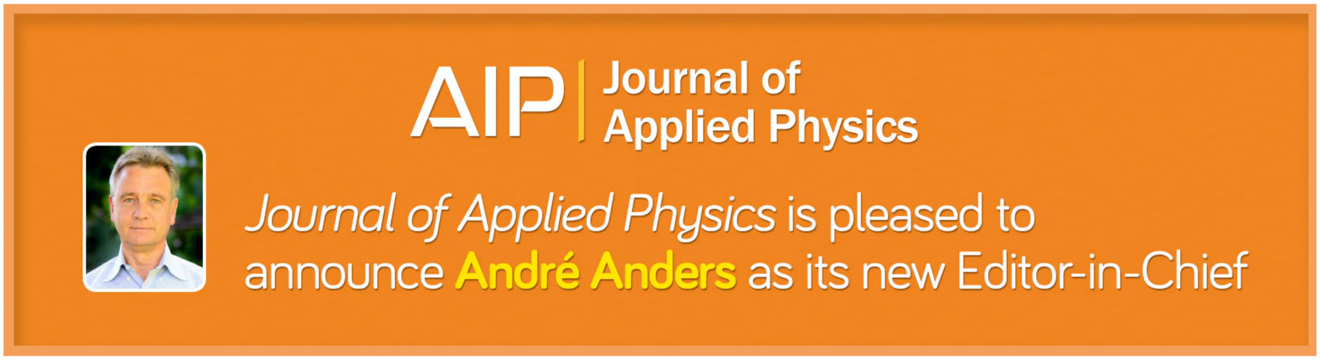
Title	Monte Carlo configuration interaction with perturbation corrections for dissociation energies of first row diatomic molecules: C-2, N-2, O-2, CO, and NO
Author(s)	Kelly, Thomas P.; Perera, Ajith; Bartlett, Rodney J.; Greer, James C.
Publication date	2014-02-28
Original citation	Kelly, T. P., Perera, A., Bartlett, R. J. and Greer, J. C. (2014) 'Monte Carlo configuration interaction with perturbation corrections for dissociation energies of first row diatomic molecules: C2, N2, O2, CO, and NO', The Journal of Chemical Physics, 140(8), pp. 084114. doi:10.1063/1.4866609
Type of publication	Article (peer-reviewed)
Link to publisher's version	http://dx.doi.org/10.1063/1.4866609 Access to the full text of the published version may require a subscription.
Rights	**© 2014, AIP Publishing. This article may be downloaded for personal use only. Any other use requires prior permission of the author and AIP Publishing. The following article appeared in J. Chem. Phys. 140, 084114 (2014); and may be found at http://dx.doi.org/10.1063/1.4866609**
Item downloaded from	http://hdl.handle.net/10468/3490

AIP | The Journal of Chemical Physics

Monte Carlo configuration interaction with perturbation corrections for dissociation energies of first row diatomic molecules: C2, N2, O2, CO, and NO

Thomas P. Kelly, Ajith Perera, Rodney J. Bartlett, and James C. Greer

Citation: The Journal of Chemical Physics **140**, 084114 (2014); doi: 10.1063/1.4866609
View online: http://dx.doi.org/10.1063/1.4866609
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/140/8?ver=pdfcov
Published by the AIP Publishing



THE JOURNAL OF CHEMICAL PHYSICS **140**, 084114 (2014)

Monte Carlo configuration interaction with perturbation corrections for dissociation energies of first row diatomic molecules: C_2, N_2, O_2, CO, and NO

Thomas P. Kelly,[1] Ajith Perera,[2] Rodney J. Bartlett,[2] and James C. Greer[1,a)]
[1]*Tyndall National Institute, University College Cork, Dyke Parade, Lee Maltings, Cork, Ireland*
[2]*Quantum Theory Project, 2234 New Physics Building #92, PO Box 118435, University of Florida at Gainesville, Gainesville, Florida 32611-8435, USA*

(Received 27 November 2013; accepted 11 February 2014; published online 28 February 2014)

Dissociation energies for the diatomic molecules C_2, N_2, O_2, CO, and NO are estimated using the Monte Carlo configuration interaction (MCCI) and augmented by a second order perturbation theory correction. The calculations are performed using the correlation consistent polarized valence "triple zeta" atomic orbital basis and resulting dissociation energies are compared to coupled cluster calculations including up to triple excitations (CCSDT) and Full Configuration Interaction Quantum Monte Carlo (FCIQMC) estimates. It is found that the MCCI method readily describes the correct behavior for dissociation for the diatomics even when capturing only a relatively small fraction (~80%) of the correlation energy. At this level only a small number of configurations, typically $O(10^3)$ from a FCI space of dimension $O(10^{14})$, are required to describe dissociation. Including the perturbation correction to the MCCI estimates, the difference in dissociation energies with respect to CCSDT ranges between 1.2 and 3.1 kcal/mol, and the difference when comparing to FCIQMC estimates narrows to between 0.5 and 1.9 kcal/mol. Discussions on MCCI's ability to recover static and dynamic correlations and on the form of correlations in the electronic configuration space are presented. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4866609]

I. INTRODUCTION

Prediction of molecular dissociation energies is a challenge for any computational method. To accurately predict dissociation energies as well as qualitatively correct dissociation curves, an even handed treatment of electron correlation across a potential energy surface is required. Post-Hartree-Fock methods and density functional theory (DFT) have different advantages and disadvantages when it comes to the theoretical treatment of dissociation and reactions. Coupled cluster methods are well suited to calculation of dissociation properties of molecules due to their size extensivity.[1] However, coupled cluster methods can encounter difficulties when treating potential energy surfaces, although unrestricted orbitals and multi-reference formalisms can eliminate some of these shortcomings. Methods such as coupled-cluster singles and doubles with perturbative triples (CCSD(T)) fail for dissociating systems. Similarly, many body perturbation methods fail for treating dissociation surfaces when a single reference description fails. Second order many body perturbation theory methods typically capture approximately 80% of the correlation energy[2] leading to relatively small differences in total energies, but still leading to relatively large errors in calculation of dissociation energies. These problems can be alleviated somewhat by introducing multi-reference versions of the theories and by using unrestricted orbitals. For both coupled cluster and many-body perturbation theory, the introduction of multi-reference determinants complicates the formalism, and in general the introduction of unrestricted orbitals can introduce spin contamination distorting the description of a potential energy surface. DFT methods can achieve dissociation energies for small molecules within typically 15 kcal/mol of experiment reflecting the overestimation of atomization energies typical of approximate DFT. To achieve more accurate results for dissociation energies requires introducing an empirical calibration to choose the best combination of approximate exchange-correlation (XC) functional and basis set.[3] DFT as commonly implemented is also confronted with the problem of using a single determinant to describe a dissociation surface. Another issue is that the XC potentials typically used in approximate DFT are symmetry independent, and can lead to incorrect results, for example, for the carbon atom. For cases where dissociation is to states which can be described by a single determinant, then a symmetry independent XC functional can be an appropriate approximation.[4] However, in cases were the states involved in dissociation are not well described by a single determinant, the combination of a single reference Ansatz and a symmetry independent XC functional introduces obstacles to a correct description of molecular dissociation.

Full configuration interaction (FCI) calculations are size extensive and size consistent, and provide the exact energy within a basis set, so therefore in principle can accurately describe potential energy surfaces and by extension dissociation energies. However, as is well known, due to the combinatorial increase in the number of determinants or spin adapted configuration state functions (CSFs) used in a CI expansion, the problem rapidly becomes computationally intractable. Hence,

---

[a)] jim.greer@tyndall.ie

CI methods rely on truncation based on excitation level ("singles," "doubles," "triples," ...) relative to a reference configuration or small set of configurations,[5] by restricting the space to which electrons may be excited,[6] or by refining the set of configurations during the diagonalization of the CI matrix.[7]

Alternative to these pre-selection schemes, methods whereby the CI expansion is randomly selected and iteratively improved have been introduced,[8–11] in a similar spirit determinant based renormalization group ideas have been applied to quantum chemistry.[12–16] In all of these methods the wave function is adaptively refined to include a relatively small number of configurations from the total CI space selected for an accurate determination of the correlated many-electron wave function and energy. Recently, similar ideas for adaptively defining the coupled cluster expansion[17] have been introduced. Diffusion Monte Carlo ideas have also recently been introduced that spawn walkers on a determinant space by randomly branching from a sample of determinants to enable importance sampling over configurations to provide stochastic calculations near the FCI[18,19] and coupled cluster[20] limits. These methods can be contrasted to multi-determinant wave functions that are selected by an iterative perturbative selection of important configurations, which are then used as an initial trial function within the fixed node approximation in a diffusion Monte Carlo calculation.[21]

In the following, dissociation energies are treated by a specific adaptive wave function method, Monte Carlo configuration interaction (MCCI).[9,22] MCCI is a technique that randomly selects electronic configurations, performs a matrix diagonalization using the Hamiltonian generated from the sampled set of configurations, retains important configurations, and discards those configurations with a low weight in the wave function, or those that make a small contribution to the energy. MCCI has been applied to the calculation of electronic spectra, charge transport in molecules, dissociation energies, multipole moments, ionization energies, and electron affinities.[23–27] In this way, the need for working with large CI vectors is avoided while a large fraction (for reasonable convergence parameters and 10–20 electrons, typically >96%) of the total correlation energy is readily obtained. The ability of the MCCI method to recover static correlation to qualitatively describe potential energy dissociation curves, as well as the ability to describe dynamic correlations for accurate estimates of dissociation energies, is studied.

II. METHODS

A. MCCI algorithm

The MCCI method is described in Refs. 8, 9, and 28 and here key features of the algorithm are highlighted. A calculation is initiated by selecting a trial vector. For all MCCI calculations presented, the TURBOMOLE[29] program with a correlation consistent polarized valence "triple zeta" (cc-pVTZ) basis[30] is used to generate a set of one- and two-electron Hartree-Fock molecular integrals. The molecular integral set is used to generate a restricted closed- or open-shell, as appropriate, Hartree-Fock determinant as a trial vector; however, the method is not in general sensitive to the initial choice. For example, the trial vector can be a linear combination of N_0 CSFs,

$$|\Psi^0\rangle = \sum_{A=1}^{N_0} c_A |\Phi_A\rangle, \tag{1}$$

where the superscript denotes the initial choice. With the choice of orthogonal molecular orbitals, only CSFs generated through single and double excitations relative to those configurations in a trial vector will result in non-zero energy matrix elements for the molecular Coulomb Hamiltonian operator. Thus, by randomly applying single and double excitations relative to the trial space, new interacting configurations can be generated. New configurations are generated from configurations in the trial vector as

$$|\Phi_B\rangle = \hat{\alpha}_r |\Phi_A\rangle, \tag{2}$$

where $\hat{\alpha}_r$ is a product of creation and annihilation operators which for a singles substitution is given as $\hat{a}_a^\dagger \hat{a}_i$, whereas for a doubles substitution it is $\hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_i \hat{a}_j$. No substitution implies multiplication by zero. The indices i, j refer to orbitals occupied in the configuration upon which the annihilation operators act, whereas the indices a, b refer to unoccupied orbitals in the CSF upon which the creation operators act. The set of new configurations B along with the CSFs A already present in the trial vector are taken together as a CI expansion for a new cycle of the adaptive sequence for improving the trial function,

$$|\Psi^{n+1}\rangle = \sum_{A=1}^{N_n} c_A |\Phi_A\rangle + \sum_{B=N_n+1}^{N_n+N_{\text{new}}} c_B |\Phi_B\rangle, \tag{3}$$

with the superscript indexing the cycle for which the trial vector is generated. The CI matrix eigenvalue problem of dimension $N_n + N_{\text{new}}$ is solved. The resulting CI coefficients and their associated configurations are retained if their magnitudes are above a specified threshold value; note configurations in the original trial vector may also be removed during this process thereby removing bias introduced by the choice of the initial trial vector. All the retained configurations are collected as a trial vector to start the next cycle. Single and double excitations are generated with respect to the configurations in the new trial vector, and the process is repeated until convergence in the energy and/or the length of the CI expansion is achieved. Repeated application of $\hat{\alpha}$ to CSFs in the trial vectors during each MCCI cycle results in a sequence of CSFs with increasing levels of excitations. After n cycles, it is possible to have reached any excitation level between 0 and $2n$ relative to the configurations in the initial trial vector. In practice, $2n \gg N$, where N is the maximum excitation in a N-electron wave function. Thus, as the configuration space is explored by the random generation of configurations, all excitation levels may be become included into the many-electron wave function *if* required. As the threshold selection for configurations is lowered, the approximation from the trial vector relative to the FCI wave function is improved. For larger values of the coefficient threshold, importance sampling is built into the algorithm as only configurations that interact with the trial vector are generated. As the trial vector at any stage of the calculation represents only a small fraction of the FCI space, the interacting configuration set defines a small sub-space of the FCI space. Using the MCCI method, a strongly interacting set of CSFs can be quickly identified, defining a "reference space" that describes static correlations and an increasing fraction of the dynamical correlations as the selection threshold for new configurations is reduced. The "reference spaces" defined in this way are much larger than those considered in conventional multi-reference approaches, for the diatomic molecules studied in the cc-pVTZ basis the dimensions vary between the order of 10^3 and 10^4 configurations.

As static correlations typically are described by low lying excitations relative to a dominant or set of few dominant configurations, the MCCI method is capable of rapidly identifying static correlations. In fact, a very useful aspect of MCCI is to determine the multi-reference nature of a wave function by sampling the CI space at a relatively large selection threshold, retaining only the leading configurations in an expansion. Dynamic correlations are described by the much larger set of configurations that weakly interact with the reference space. If essentially all of the correlation energy is sought from a calculation and the dynamic correlations are spread throughout a large fraction of the CI space, Monte Carlo techniques become increasingly inefficient. However, for many cases of interest, the correlation energy can be described to within a high level of accuracy using only a small fraction of the FCI space consisting primarily of singles, doubles, triples, and quadruple (SDTQ) excitations, and indeed only requiring a small subset of the SDTQ excitations. It can be argued for larger problems that Monte Carlo techniques become the only pragmatic approach to identifying important configurations in explicit determinant or CSF descriptions of the many-electron wave function.

The key parameter for selecting the accuracy of an MCCI calculation is the criterion for retaining CI expansion configurations with coefficients $|c_A| \geq \texttt{threshold}$ into the new MCCI cycle as part of the next trial vector. It is intuitive that for small values of the coefficients $|c_A|$ the corresponding CSF may be disregarded. However, the amount of correlation that can be obtained for a fixed value of the selection threshold varies for different molecules, or even for the same molecule in different states. Hence, for determining the fraction of the correlation energy obtained from an MCCI calculation, the selection threshold is an ambiguous measure. Intermediate normalization can be introduced where the leading reference determinant is normalized as $\langle \Phi_0 | \Phi_0 \rangle = 1$, or by introducing a multi-reference variant of intermediate normalization for systems with strong static correlations. In this way, the weight of a configuration can be measured against the weight of the reference space in the total wave function introducing some degree of transferability to the selection criterion. However, this remains a somewhat *ad hoc* procedure and also suffers in that it does not provide an estimate for the magnitude of the neglected electron correlations.

To obtain numbers relevant to, for example, spectroscopy and thermochemistry, difference energies are required. The difference in FCI energies between systems X and Y, or for the same molecule in different states X and Y, is

$$\Delta E = E_X - E_Y. \tag{4}$$

This may be re-expressed as

$$\Delta E = \tilde{E}_X - \tilde{E}_Y + \delta E_X - \delta E_Y, \tag{5}$$

where $\tilde{E}_{X/Y}$ are the approximations to the FCI energy and $\delta E_{X/Y}$ are the errors in the approximate energies. CI energies are variational ensuring that $\delta E_{X/Y} \geq 0$, but difference between the error terms may be positive or negative; however in all cases the two variational errors will at least partially cancel. Hence, if an approximate CI can be performed and the magnitude of the error for the correlation energy between systems or states of a system are approximately equal or that the error magnitudes are sufficiently small that the difference energy is negligible, the approximation to the energy difference

$$\Delta \tilde{E} = \tilde{E}_X - \tilde{E}_Y \tag{6}$$

will be a good approximation to the FCI energy difference Eq. (5). Error cancellation has been used to advantage to study electronic excitations in atoms and small molecules.[23,32] In those studies, the magnitude for the threshold for retaining configurations in the MCCI procedure was allowed to remain relatively large, yet accurate predictions for vertical excitation energies were achieved. The accuracy achieved in the calculations indicates that the error in the correlation energy between the ground and excited states is of a similar magnitude.

In relation to the present study, dissociation energies arise from the difference in energy at the molecular equilibrium and dissociated geometries. In contrast with the calculation of electronic excitations, the difference in correlation energies as well as the error in predicting the correlations energies is expected to vary substantially between the equilibrium and dissociated states, as it is well known the character of the wave function varies substantially as the molecule dissociates. Hence, an accurate prediction of neglected correlations δE_X, δE_Y is required for an accurate estimate of dissociation energies when computed using the MCCI method. In the following, MCCI will be used to determine dissociation curves and, with the aid of a perturbation estimates to the error terms to be discussed next, total energies are obtained to an accuracy that enables prediction of dissociation energies to near chemical accuracy (∼1–2 kcal/mol).

B. MCCI + A_K estimates

The idea of selectively including configurations into a multi-reference CI singles and doubles calculation (MR-CI(SD)) and then estimating the effect of the neglected configurations dates back to the work of Buenker and Peyerimhoff.[31] In their approach, the energy of a full MR-CI(SD) calculation is estimated. In the following, using MCCI to define a "reference space" that captures both static correlations and the leading dynamic correlations, a similar approach is applied to estimate the effect of the neglected dynamic correlations. Defining $|\Psi\rangle$ as a final MCCI wave function with energy E, it follows that for all CSFs, $|\Phi_K\rangle$, that interact with the MCCI wave function

$$\Psi^{(1)} = \sum_K \frac{|\Phi_K\rangle\langle\Phi_K|\hat{H}|\Psi\rangle}{E - E_K}, \tag{7}$$

$$E^{(2)} = \langle\Psi|\hat{H}|\Psi^{(1)}\rangle, \tag{8}$$

where $E_K = \langle\Phi_K|\hat{H}|\Phi_K\rangle$. This is the idea behind the A_K method[33] which uses second order perturbation estimates of the energy to determine important configurations for inclusion in a CI calculation. Harrison[34] uses perturbation theory ideas by generating all singles and doubles with respect to a reference space, identifying the important configurations using a perturbation estimate, adaptively expanding the reference space, performing a CI calculation on this larger CI vector, and iterating to convergence. Within the context of MCCI calculations, the perturbative estimates have been used to pre-screen the randomly chosen configurations to be included within the CI vector prior to a matrix diagonalization step.[35] Recently Coe and Paterson[26] have introduced a hybrid method which begins by first converging an MCCI calculation to a given configuration threshold selection value, then continues by adaptively refining the CI vector in a scheme similar to Harrison's algorithm.

In the following, a variant on these ideas will be introduced to augment MCCI estimates of total correlation energies using a single perturbative estimate. In this approach, the MCCI algorithm is used to determine the important configurations for relatively large reference wave functions, and for which it becomes relatively time consuming to generate all single and double excitations. A MCCI vector is obtained to a desired convergence level and then resulting CSFs are ordered. The CI vector is truncated to a fixed number of dominant CSFs. The resulting CI vector is re-diagonalized and taken as a reference wave function, and all single and double excitations with respect to the truncated vector are generated. The individual energy contributions from each of the singly and doubly excited configurations with respect to the reference wave function are then summed to determine a perturbative correction. A particularly convenient aspect of this approach is that the estimates are trivially parallelizable allowing for a perturbation correction in conjunction with the use of MCCI to identify large reference spaces.

III. RESULTS

A. Static correlations and molecular dissociation

Before proceeding to the estimation of dissociation energies, the ability to recover static correlations in producing approximations to molecular dissociation curves using MCCI is considered. The potential energy as a function of the dissociation coordinate is given in Fig. 1 for the homonuclear diatomics C_2 and N_2. The dissociation curves were generated by performing MCCI calculations at the Hartree-Fock equilibrium bond length R_0 and for 2×, 3×, and 4× the equilibrium bond length. The CSFs obtained for each bond length are merged into a single CI vector and then used to determine the potential energy as a function of the dissociation coordinate. For this procedure, a self-consistent set of molecular orbitals is re-calculated for each value of the bond length.

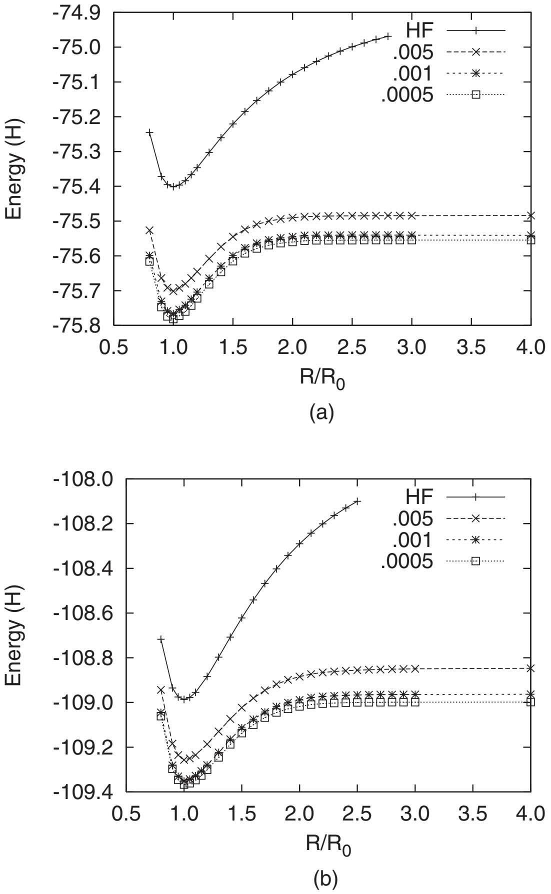


FIG. 1. MCCI dissociation energy curves. (a) Dissociation curve for C_2 in the singlet ground state determined by MCCI compared to the Hartree-Fock approximation. The potential energy curve is calculated from the CI vector obtained from merging the MCCI vectors obtained from calculations at 1×, 2×, 3×, and 4× the equilibrium bond length at a fixed selection threshold. This procedure is repeated for coefficient selection thresholds of 0.005 yielding 1021 CSFs, 0.001 yielding 7924 CSFs, and 0.0005 yielding 17 967 CSFs. At a selection threshold of 10^{-3}, the 1×, 2×, 3×, and 4× R_0 bond length calculations contributed 2612, 743, 284, and 257 unique CSFs, respectively, of the total of 7924 CSFs. (b) Dissociation curve for the singlet ground state of N_2 calculated using the CI vector obtained from merging MCCI vectors obtained at 1×, 2×, 3×, and 4× R_0 and repeated for coefficient threshold values of 0.005 yielding 1055 CSFs, 0.001 yielding 10 680 CSFs, 0.0005 yielding 19 700 CSFs. At a selection threshold of 10^{-3}, the 1×, 2×, 3×, and 4× R_0 bond length calculations contributed 1741, 1836, 516, and 611 unique CSFs, respectively, from a total of 10 680 CSFs. R_0 in both cases is taken to be the Hartree-Fock equilibrium geometry (2.345 a.u. for C_2 and 2.017 a.u. for N_2).

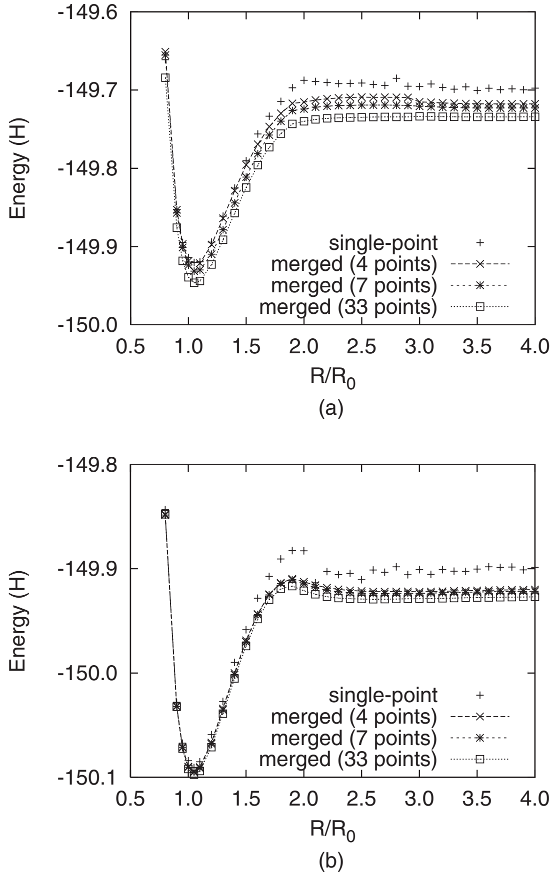
In Fig. 1(a), the ground state energy at equilibrium to dissociation in the singlet state is plotted for the C_2 in the Hartree-Fock approximation, and for MCCI with a selection threshold $|c_A| > 5 \times 10^{-3}$, $|c_A| > 10^{-3}$, and for $|c_A| > 5$

$\times\ 10^{-4}$. As is well known, the Hartree-Fock approximation is incapable of correctly describing dissociation. On the other hand, capturing only a small amount of the electronic correlation energy with the selection threshold of $|c_A| > 5 \times 10^{-3}$ yielding only 2042 configurations enables a reasonable description of the behavior of the energy between the equilibrium and dissociated geometries. Similar results for N_2 are obtained and displayed in Fig. 1(b). In analogy to the nonparallelity error (NPE) which is a measure of the maximum deviation between an approximate potential energy surface and the FCI surface independent of any constant shift between them, the deviation Δ between the curve calculated with a selection thresholds of 10^{-3} and 5×10^{-4} defined as

$$\Delta = \max\left|E_R^{\text{ref}} - E_R\right| - \min\left|E_R^{\text{ref}} - E_R\right|, \tag{9}$$

where E_R^{ref} and E_R are the energies of the reference calculation and the energy of the approximation being compared, respectively, at identical bond lengths R and allowing the bond lengths to vary between R_0 and $4R_0$. Δ is less than 3 kcal/mol in the case of C_2 and is 12 kcal/mol for N_2.

Shown in Fig. 2(a) is a comparison, using molecular oxygen as an example, of how the number of independent MCCI calculations used to build a merged vector affects the shape of a molecular dissociation curve. As before, the MCCI vectors from different bond lengths are merged resulting in a single CI vector consisting of all CSFs obtained from the calculations at the different bond lengths. Merged CI vectors are constructed using either 4, 7, or 33 equally spaced points between R_0 and $4R_0$ obtained from MCCI calculations with a selection threshold of 5×10^{-3}. The merged vector obtained from 4 points contains a total of 1266 CSFs, while the merged vector obtained from 33 distinct bond lengths contains 2248 CSFs. The overall shape of the dissociation curve remains similar for all curves depicted, however to achieve this result it is important that the selected bond lengths are chosen to uniformly span the entire range from equilibrium to dissociation. Figure 2(b) shows curves constructed from the same vectors as in (a) with addition of second order perturbation theory corrections. When including the perturbation corrections, the shape of the curve is less sensitive to the number of calculations at distinct bond lengths that are used to form the merged vector. In Fig. 3, similar data are shown for the quintet state of C_2 which at dissociation is the ground state of C + C. Dissociation curves were constructed by merging the MCCI vectors obtained from the bond lengths of 1×, 1.5×, 2×, 2.5×, and 3× R_0. In this case, the Hartree-Fock approximation incorrectly predicts a bound state, including only low levels of correlation from MCCI calculations with $|c_A| > 5 \times 10^{-3}$ resulting in a merged vector length of only 1350 CSFs is found to behave similarly to a description of the dissociation generated from MCCI calculations with a selection threshold of $|c_A| > 10^{-3}$ with a merged CI vector of 8305 CSFs. As with the singlet state, it is important to include bond lengths spanning from equilibrium to dissociation, though geometries closer to the equilibrium geometry contribute the largest number of unique CSFs, i.e., CSFs that are not included by the MCCI calculations at the other bond lengths. For the $|c_A| > 10^{-3}$ calculation, the R_0 geometry contributes 1659 unique CSFs to the CI vector, while at the $3 \times R_0$ geometry 453 unique CSFs are found. The calculations are consistent with the well known fact that static correlations to a large extent govern the qualitative behavior for molecular



(a)

(b)

FIG. 2. Triplet ground state O_2 calculation obtained from (a) CI vectors and (b) CI vectors plus second order perturbation corrections obtained by merging MCCI vectors calculated at 4, 7, and 33 points along the dissociation curve. All MCCI vectors obtained using a coefficient selection threshold of 5 $\times\ 10^{-3}$. Merged vector lengths range from 1266 CSFs to 2248 CSFs. Single-point calculations are also shown for comparison. R_0 is the Hartree-Fock equilibrium geometry (2.177 a.u.), whereas the minima including correlations is found at approximately 1.05 × the Hartree-Fock minimum.

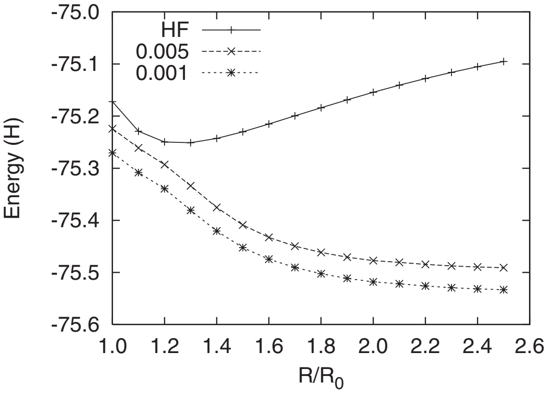


FIG. 3. Quintet state of C_2 which is the ground state at dissociation C + C. The Hartree-Fock approximation incorrectly predicts a bound state at shorter bond lengths, while MCCI provides qualitatively the correct dissociation in the quintet state using only 1350 CSFs obtained at a coefficient selection threshold of 0.005, representing a small fraction of the full CI space.

dissociation, and we have demonstrated that the configurations describing static correlations are readily identified from a Monte Carlo search of the CI space even at a relatively large selection threshold. The calculations yield results that are computationally efficient insofar as it is only required to diagonalize small matrices, and whether a molecule is described by a single- or multi-reference expansion poses no particular challenge to the method.

B. Capturing dynamic correlations

Dynamic correlations must be treated in an even handed manner across the dissociation pathway to yield energy differences that are accurate and in contrast to static correlations, a much larger number of configurations are typically required to describe dynamic correlations. As previously discussed, it is required to estimate the error magnitudes in Eq. (5) between two total energy calculation at different geometries and for different spin states are comparable for accurate estimation of dissociation energies. For the MCCI calculations, C_2, N_2, and CO are calculated as singlets at their equilibrium bond lengths, NO as a doublet, and O_2 as a triplet. The dissociated limit is taken to be eight times the equilibrium bond length and the calculations are performed in the energetically favorable spin states with C+C, O+O, and C+O quintets, N+O a hextet, and N+N a septet. The lowest selection threshold chosen for the calculations is 2×10^{-4}. Based on Full Configuration Interaction Quantum Monte Carlo (FCIQMC) estimates for the full configuration interaction energies presented in Ref. 19, it is found that the MCCI energies presented in Table I recover over 93% of the total correlation energy for the molecules at their equilibrium geometries, and over 97% of the total correlation energy for dissociated geometries; it should be noted that even at this relatively high selection threshold the total energies are comparable to multi-reference CI(SD) calculations, although this result is obtained from a much smaller number of configurations. The NO molecule at equilibrium has the largest percentage of correlation energy neglected by MCCI of 6.7% at a selection threshold of 2×10^{-4}, suggesting a large percentage of dynamic correlation must be captured in this case to approach the FCI limit. Also in Table I is a comparison to CCSD total and dissociation energies. All coupled cluster calculations presented in this work were carried out using the ACES[36] program. As mentioned previously, the selection coefficient threshold is not transferable in the sense that differing amounts of the correlation energy are obtained between equilibrium and dissociated geometries. Table I highlights this limitation where it is found that the estimates for the dissociation energies obtained from MCCI at a fixed selection threshold do not perform consistently compared to the size extensive CCSD predictions. For small enough selection thresholds, the calculations for both the equilibrium and dissociated geometries will approach FCI results with size extensivity restored, but for a reasonably large selection threshold this will not be the case. The convergence for the NO molecule at the equilibrium and dissociated geometries is shown in Fig. 4 as a function of the number of optimal CSFs included in the MCCI vector. Clear to see from the figure is that the dissociated geometry is much easier to converge toward FCI limits with a smaller set of CSFs, whereas the dynamic correlations at equilibrium are more difficult to capture. It can be inferred from Table I and Fig. 4 that it is difficult to balance the magnitude of the neglected correlations to cancel the error between equilibrium and dissociated geometries using a single selection threshold. Another feature of the convergence in the correlation energy as shown in Fig. 4 is the fact that there are two distinct regions that exist beyond the range of configurations that would normally be classified as static correlations. The two regions are defined

TABLE I. A comparison of MCCI energies (converged with a coefficient threshold of 2×10^{-4}) with CCSD. Absolute energies are reported in hartrees. Dissociation energies are reported in kcal/mol. With the exception of C_2 and N_2, MCCI estimates are comparable to CCSD. The equilibrium bond lengths R_0 are taken as the CCSDT geometries found in Ref. 37 and are 2.3632 bohrs for C_2, 2.0845 bohrs for N_2, 2.898 bohrs for O_2, 2.1460 bohrs for CO, and 2.1853 bohrs for NO. The dissociated bond lengths are taken to be $8 \times R_0$.

	E_{Eq}	N_{Eq}	E_{Diss}	N_{Diss}	D_0	ΔD_0
C_2						
MCCI	− 75.7740	30 203	− 75.5536	20 960	138.79	
CCSD	− 75.7496		− 75.5571		121.27	−17.51
N_2						
MCCI	− 109.3578	24 139	− 109.0183	14 157	213.91	
CCSD	− 109.3552		− 109.0245		208.34	−5.57
O_2						
MCCI	− 150.0989	29 213	− 149.9256	14 239	109.17	
CCSD	− 150.1096		− 149.9418		105.71	−3.46
CO						
MCCI	− 113.1329	25 362	− 112.7472	9 832	242.94	
CCSD	− 113.1384		− 112.7494		245.07	2.13
NO						
MCCI	− 129.6899	36 674	− 129.4799	9 970	132.27	
CCSD	− 129.6978		− 129.4831		135.26	2.99

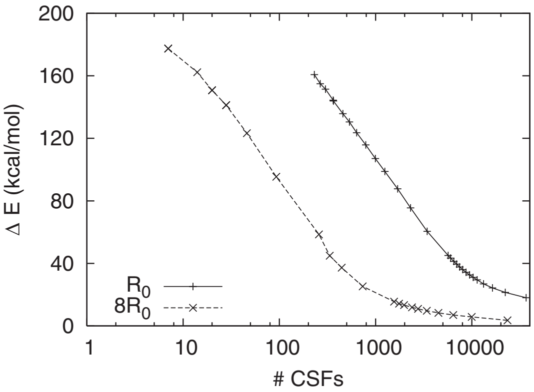


FIG. 4. Difference between FCIQMC and MCCI total energies for the NO molecule as a function of the number of CSFs in the MCCI vectors at R_0 and $8R_0$ bond lengths. The curves represent the convergence of the MCCI calculations as the coefficient selection threshold is reduced and reveal that beyond the range of what is normally considered to be static correlations that there are two distinct regimes of dynamic correlations as indicated by the changing slopes of the curves at larger number of CSFs. Note the energies are given in kcal/mol for subsequent comparison to reporting of dissociation energies.

by different slopes in the convergence curves; for purposes of discussion, the region with the steeper slope occurring at a smaller number of CSFs is labeled type 1 correlations, and the region with slower convergence occurring at larger number of CSFs is labeled type 2 correlations. For the dissociated geometry N+O, the transition between the two regions occurs in the vicinity of 1000 CSFs, whereas for the equilibrium geometry this transition is seen to occur for >10 000 CSFs. For the equilibrium state, a much larger absolute energy error must be recovered by type 2 correlations relative to the equilibrium state.

In Fig. 5, the energy contributions are presented in matrix form, with each pixel representing an average over 100 × 100 energy contribution magnitudes and the leading 12 000 CSFs from each of the MCCI calculation have been chosen for depiction of the CI space. Each energy contribution is defined as

$$E_{IJ} = c_I^* H_{IJ} c_J / \vec{c}^{\,\dagger} \underline{S} \vec{c}, \tag{10}$$

with H_{IJ} the energy matrix element between configurations I and J. The choice of 12 000 CSFs will be motivated later in this section when the size of the CI vector from which perturbative estimates are generated is considered to achieve a given accuracy in the dissociation energies. The lighter gray colors correspond to lower values of the energy contributions and the darker colors represent larger energy contributions. The off-diagonal elements in the figure represent the magnitude of the interactions between CSFs and hence describe the nature of the molecular correlations. Depicted within Fig. 5 are energy contributions for the N_2 molecule and the O_2 molecule and by comparison between the two, it can be concluded that the set of configurations with strong interactions is much larger for the case of the O_2 molecule with respect to the N_2 molecule. A similar depiction for the NO molecule is shown in Fig. 6 but in this case comparing the correlations between the equilibrium and dissociated geometries. In this case it is clear that the set of equilibrium configurations that are highly interacting is much larger than for the dissociated state. This then gives insight into the two regimes for the dynamical correlations. Type 1 dynamical correlations can be found relatively easily by the MCCI algorithm, but type 2 dynamical correlations are much more weakly interacting among themselves

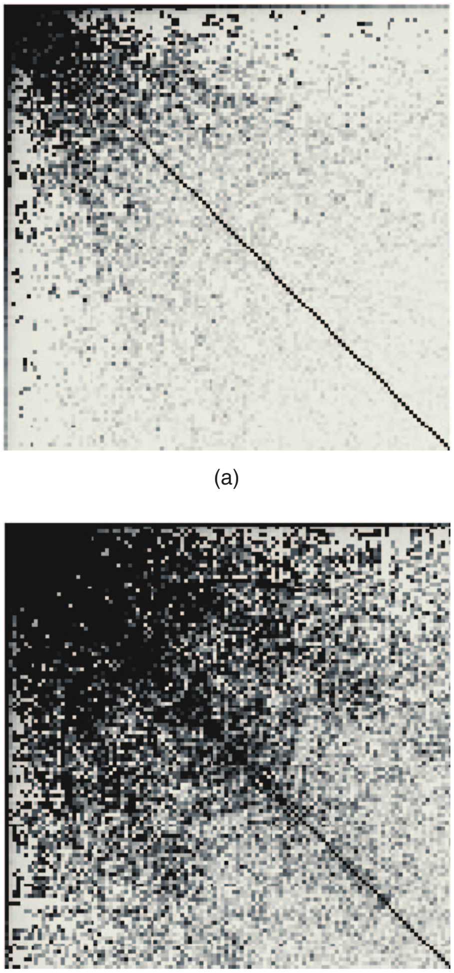


FIG. 5. Depiction of MCCI energy contributions constructed from CI vectors of length 12 000 for (a) the N_2 equilibrium geometry, (b) the O_2 equilibrium geometry. The gray scale indicates a linear scale for the magnitude of each energy contribution with the lighter shading corresponding to lower values.

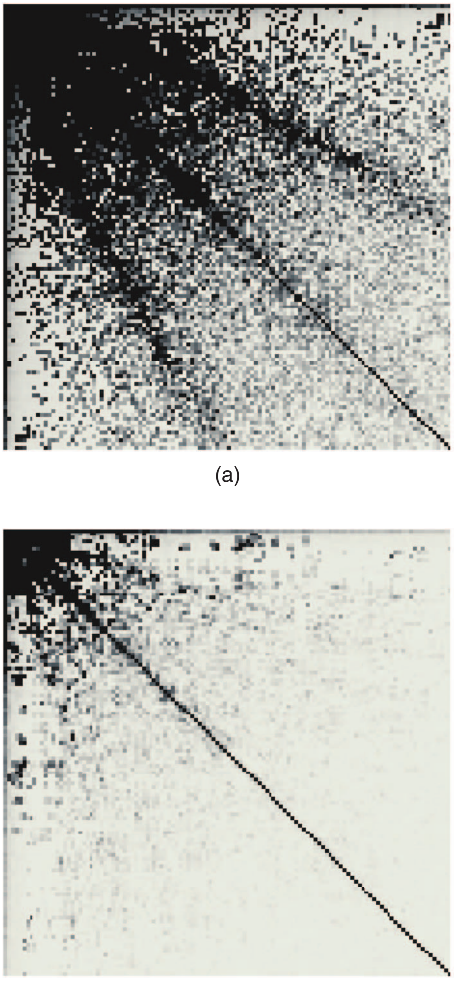


FIG. 6. Depiction of MCCI energy contributions constructed from CI vectors of length 12 000 for (a) the NO equilibrium geometry, (b) the N O dissociated state. The gray scale indicates a linear scale for the magnitude of each energy contribution with the lighter shading corresponding to lower values.

TABLE II. Total energies and dissociation energies estimates in the cc-pVTZ basis. A reference space of 12 000 CSFs was used in the CI calculations for the reference spaces. N_{PT2} is the total number of unique single and double excitations relative to the reference space and E_{PT2} is the perturbation estimates from these configurations. All energies are reported in hartree with the exception of dissociation energies, which are reported in kcal/mol. The homonuclear molecules overestimate dissociation energy, while the heteronuclear molecules underestimate dissociation energy with respect to both CCSDT and FCIQMC. The equilibrium bond lengths R_0 are taken as the CCSDT geometries found in Ref. 37 and are 2.3632 bohrs for C_2, 2.0845 bohrs for N_2, 2.898 bohrs for O_2, 2.1460 bohrs for CO, and 2.1853 bohrs for NO. The dissociated bond lengths are taken to be 8 × R_0. Equilibrium bond lengths for the FCIQMC calculations are taken from Ref. 38 (differences between the CCSDT and experimental geometries vary between >0.001 and <0.01 bohr).

Equilibrium bond length

	$E_{MCCI+PT2}$	N_{PT2}	E_{PT2}	ΔE (CCSDT)	ΔE (FCIQMC)
C_2	−75.7837	10 539 816	−0.0182	−0.0026	0.0013
N_2	−109.3739	24 298 236	−0.0217	−0.0002	0.0015
O_2	−150.1216	40 661 401	−0.0380	0.0071	0.0089
CO	−113.1495	40 688 511	−0.0236	0.0062	0.0069
NO	−129.7127	53 840 989	−0.0420	0.0052	0.0058

Dissociated bond length

	$E_{MCCI+PT2}$	N_{PT2}	E_{PT2}	ΔE (CCSDT)	ΔE (FCIQMC)
C + C	−75.5600	12 945 063	−0.0098	0.0023	0.0023
N + N	−109.0266	23 228 804	−0.0089	−0.0032	0.0028
O + O	−149.9393	42 276 847	−0.0147	0.0090	0.0095
C + O	−112.7530	54 090 448	−0.0074	0.0023	0.0025
N + O	−129.4866	67 369 019	−0.0056	0.0023	0.0024

Dissociation energies

	D_0(MCCI + PT2)	ΔD_0 (CCSDT)	ΔD_0 (FCIQMC)
C_2	140.89	−3.08	−1.20
N_2	218.76	−2.142	−1.90
O_2	114.84	−1.19	−0.49
CO	249.80	2.45	1.86
NO	142.40	1.81	1.5

and interact back with only a small number of configurations from the reference space (static correlations) and type 1 dynamical correlations. The latter set of configurations capture a much smaller amount of correlation energy, but an estimate of their contributions is necessary to obtain accurate estimates for dissociation energies.

To improve the estimates of dissociation energies, all single and double excitations of a reference CI vector obtained from MCCI are generated and their contribution to the total energy is estimated using the procedure outlined in Sec. II B. Table II compares these perturbation corrected MCCI energies for all the molecules considered. In Fig. 7, the convergence in the dissociation energies is studied versus the size of the MCCI-selected "reference space." It is found that to obtain dissociation energies within 2 kcal/mol of FCIQMC estimates for C_2, N_2, O_2, CO, and NO that a reference space size of approximately 12 000 CSFs should be chosen. The calculations are performed by extracting the top 12 000 most significant CSFs obtained from the MCCI vectors calculated with a coefficient selection threshold of 2×10^{-4} for equilibrium geometries and 10^{-4} for dissociated geometries. The CI matrix for this truncated vector is re-diagonalized and used as the reference wave function for the perturbation estimates. The perturbation estimate introduces error estimates for the neglected correlations in Eq. (5) with the resulting dissociation energy agreeing with FCIQMC to within ∼2 kcal/mol, and for the C_2 and O_2 molecules the differences are within 0.8 kcal/mol.

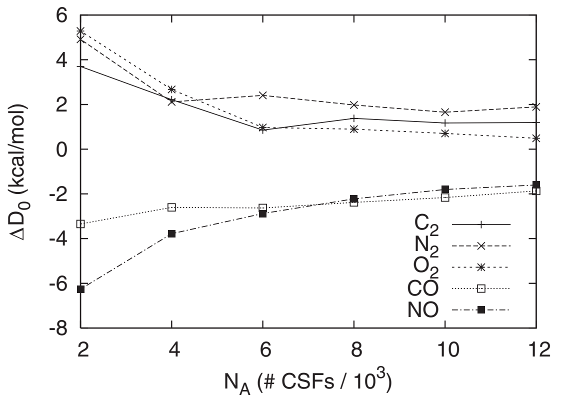


FIG. 7. Dissociation energy estimates from MCCI + perturbation correction converge with increasing reference space N_A. Dissociation energy is overestimated for the homonuclear cases, and underestimated for the heteronuclear cases, with respect to the FCIQMC estimates.

The procedure may now be approximately described as follows. The MCCI calculations are used to define static correlations and the important singles and doubles interacting back to the static correlations, these latter terms as generated can be equated to the type 1 dynamic correlations. It should be noted that with the use of the MCCI algorithm, the type 1 correlations are obtained by repeated diagonalization of matrices that are only a small fraction of the dimension of the full CI(SD) or MR-CI(SD) matrices that would otherwise be required. Relative to the static and type 1 correlations, all singles and doubles are considered using the perturbation estimate. As can be seen in Fig. 4, the type 2 correlations contributions are small relative to type 1 correlations as reflected in the change of the slope in Fig. 4 at larger number of CSFs and there is a much larger number of the type 2 configurations. Thus, these configurations are much more difficult to sample using Monte Carlo techniques as they are effectively distributed over the entire space defined by single and double substitutions with respect to the configurations describing the static and type 1 correlations. However, as the perturbation estimate is trivial to parallelize, their contribution to the "reference space" consisting of the static and type 1 correlations can be readily generated.

IV. CONCLUSIONS

MCCI quickly generates compact wave functions capable of capturing a large percentage of the correlation energy and generates qualitatively correct potential energy surfaces. The robust treatment of multi-reference wave functions by MCCI reveals the algorithm it is well suited for capturing static correlations. Examining the convergence of the MCCI calculations with respect to vector length indicates two distinct regimes for the description of dynamical correlations, which we have labeled type 1 and 2 dynamical correlations. Using the MCCI algorithm to determine the "reference space" consisting of static and type 1 dynamical correlations, augmented by a second order perturbation correction for estimating type 2 dynamical correlations leads to highly accurate estimates of dissociation energies to within 2 kcal/mol of full configuration interaction energies estimated by the FCIQMC method.

The calculations reveal that to obtain accurate dissociation energies in the range of 10^7 to 7×10^7 CSFs for the molecular set C_2, N_2, O_2, CO, and NO in the cc-pVTZ basis are required. This is readily achieved by considering convergence of the dissociation energies as a function of the "reference space" consisting of static and type 1 correlations and applying relative to this space the perturbation estimate to gauge the effect of neglected correlations. Following this approach, it is found that in the cc-pVTZ basis that approximately 12 000 configurations are sufficient to define the "reference space" needed to ensure dissociation energy estimates near chemical accuracy for the molecular test set. It has been demonstrated that the configurations for these large "reference spaces" can be found using the MCCI algorithm without the requirement to diagonalize large, within the context of CI calculations, matrices although with the introduction of the requirement to perform repeated diagonalizations of matrices on the order of the number of static contributions plus leading type 1 correlations. The sampling of the CI space during the Monte Carlo iterations can be performed in parallel resulting in efficient searches for the static and type 1 correlations. From the perturbation estimate, it is found that the neglected correlations are larger for the equilibrium geometries (0.029 hartree on average) than for dissociated geometries (0.009 hartree on average). Hence achieving the error cancellation as described in Eq. (6) is difficult for calculation of dissociation energies, and the perturbation corrections are needed to provide an accurate estimate to the FCI energy. In the study presented, the error estimates have been tested by comparing against benchmark FCIQMC calculations. However, Fig. 4 identifies a means for defining the minimum dimension for a "reference space" needed to achieve near FCI estimates of the total energies. To capture static and type 1 correlations, the "reference space" should be large enough to include all configurations up to the bend seen in Fig. 4 of the convergence of the energy with number of CSFs for the equilibrium geometry. As the correlations in the dissociated geometry are easier to capture, the dimension of the "reference space" from the equilibrium geometry should be chosen for performing the perturbation estimates for both the equilibrium and dissociated geometries to ensure a balanced treatment for the perturbation estimates. As the perturbation estimates consist of trivially parallelizable calculations, efficient estimates of dissociation energies with the method are readily attainable from the resulting total energy estimates that are found to be within a few kcal/mol of FCI limits.

ACKNOWLEDGMENTS

This work has been made possible by a Principal Investigator grant to JCG from Science Foundation Ireland.

[1] R. J. Bartlett and M. Musiał, Rev. Mod. Phys. **79**, 291 (2007).
[2] F. Jensen, *Introduction to Computational Chemistry* (John Wiley and Sons, Chichester, 2006).
[3] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[4] U. von Barth, Phys. Rev. A **20**, 1693 (1979).
[5] N. Ben-Amora, S. Evangelisti, D. Maynau, and E. Rossi, Chem. Phys. Lett. **288**, 348 (1998).
[6] E. Rossi, G. L. Bendazzoli, S. Evangelisti, and D. Maynau, Chem. Phys. Lett. **310**, 530 (1999).
[7] A. O. Mitrushenkov, Chem. Phys. Lett. **217**, 559 (1994).
[8] J. C. Greer, J. Chem. Phys. **103**, 1821 (1995).
[9] J. C. Greer, J. Comput. Phys. **146**, 181 (1998).
[10] M. C. Troparevsky and A. Franceschetti, J. Phys. Condens. Matter **20**, 055211 (2008).
[11] M. Sambataro, D. Gambacurta, and L. Lo Monaco, Phys. Rev. B **83**, 045102 (2011).
[12] G. Fano, F. Ortolani, and L. Ziosi, J. Chem. Phys. **108**, 9246 (1998).
[13] S. R. White and R. L. Martin, J. Chem. Phys. **110**, 4127 (1999).
[14] G. K. L. Chan and M. Head-Gordon, J. Chem. Phys. **116**, 4462 (2002).
[15] K. H. Marti, I. M. Ondik, G. Moritz, and M. Reiher, J. Chem. Phys. **128**, 014104 (2008).
[16] K. Boguslawski, K. H. Marti, and M. Reiher, J. Chem. Phys. **134**, 224101 (2011).
[17] D. I. Lyakh and R. J. Bartlett, J. Chem. Phys. **133**, 244112 (2010).
[18] G. H. Booth, D. Cleland, A. J. Thom, and A. Alavi, J. Chem. Phys. **135**, 084104 (2011).
[19] D. Cleland, G. H. Booth, C. Overy, and A. Alavi, J. Chem. Theory Comput. **8**, 4138 (2012).
[20] A. J. W. Thom, Phys. Rev. Lett. **105**, 263004 (2010).
[21] E. Giner, A. Scemama, and M. Caffarel, Can. J. Chem. **91**, 879 (2013).
[22] J. C. Greer, J. Chem. Phys. **103**, 7996 (1995).
[23] W. Győffry, R. J. Bartlett, and J. C. Greer, J. Chem. Phys. **129**, 064103 (2008).
[24] M. Szepieniec, I. Yeriskin, and J. C. Greer, J. Chem. Phys. **138**, 144105 (2013).
[25] J. P. Coe, D. J. Taylor, and M. J. Paterson, J. Comput. Chem. **34**, 1083 (2013).
[26] J. P. Coe and M. J. Paterson, J. Chem. Phys. **137**, 204108 (2012).
[27] J. P. Coe, D. J. Taylor, and M. J. Paterson, J. Chem. Phys. **137**, 194111 (2012).
[28] L. Tong, M. Nolan, T. Cheng, and J. C. Greer, Comput. Phys. Commun. **131**, 142 (2000).
[29] TURBOMOLE V6.4 2012, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007.

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

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

[32] J. A. Larsson, L. Tong, T. Cheng, M. Nolan, and J. C. Greer, J. Chem. Phys. **114**, 15 (2001).

[33] Z. Gershgorn and I. Shavitt, Int. J. Quantum Chem. **2**, 751 (1968).

[34] R. Harrison, J. Chem. Phys. **94**, 5021 (1991).

[35] W. Győrffy, Ph.D. thesis, University College Cork, 2007.

[36] ACES II is a program product of the Quantum Theory Project, University of Florida. Authors: J. F. Stanton, J. Gauss, J. D. Watts, M. Nooijen, N. Oliphant, S. A. Perera, P. G. Szalay, W. J. Lauderdale, S. R. Gwaltney, S. Beck, A. Balková, D. E. Bernholdt, K.-K. Baeck, P. Rozyczko, H. Sekino, C. Hober, and R. J. Bartlett. Integral packages included are VMOL (J. Almløf and P. R. Taylor); VPROPS (P. Taylor); ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, J. Olsen, and P. R. Taylor).

[37] D. Feller and J. A. Sordo, J. Chem. Phys. **113**, 485 (2000).

[38] K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure: Constants of Diatomic Molecules* (Van Nostrand Reinhold, New York, 1979), pp. 112–490.