Stochastic Coupled Cluster Theory

Alex J. W. Thom*

Department of Chemistry, Imperial College London, London SW7 2AZ, United Kingdom and University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, United Kingdom

(Received 14 September 2010; published 28 December 2010)

We describe a stochastic coupled cluster theory which represents excitation amplitudes as discrete excitors in the space of excitation amplitudes. Reexpressing the coupled cluster (CC) equations as the dynamics of excitors in this space, we show that a simple set of rules suffices to evolve a distribution of excitors to sample the CC solution and correctly evaluate the CC energy. These rules are not truncation specific and this method can calculate CC solutions to an arbitrary level of truncation. We present results of calculation on the neon atom, and nitrogen and water molecules showing the ability to recover both truncated and full CC results.

DOI: 10.1103/PhysRevLett.105.263004 PACS numbers: 31.15.bw, 31.15.V−

Coupled cluster (CC) theory has been said to be the gold standard of electronic structure methods, and is routinely used in many branches of physics [1–3]. With the addition of perturbative corrections, CCSD(T) can routinely produce energies well within ''chemical accuracy'' [4], of 1 kcal/mol. It has the desirable property that truncations of the theory are size extensive and rapidly converge with increasing truncation level to give spectroscopic quality results. Despite this, the theory has a number of drawbacks. Its scaling (starting from sixth order with system size) means that large computational resources are required for relatively small systems, though local approximations have had some success in countering this. Also, although in principle systematically improvable with truncation level, this vastly worsens the scaling and the complexity of the equations, and so high-order truncations of coupled cluster theory are only available in specialist codes.

In this Letter, we turn to stochastic methods, which have been shown able to provide favorable scalings for correlation energies [5,6], and propose a radically different formulation of coupled cluster theory, inspired by the full configuration interaction quantum Monte Carlo method of Alavi *et al.* [7–9] (which, along with its use in a truncated form we shall describe CIQMC). In this new formulation, instead of sampling determinants, we sample the space of excitation amplitudes by a set of ''excitors'' whose population dynamics converge upon the ground state coupled cluster solution.

We begin with a brief overview of the CIQMC method in a novel exposition which clarifies the derivation of the CC formulation. Within a given basis, the complete solution to the many-electron Schrödinger equation may be calculated by expressing the wave function as a sum of all possible components of the Hilbert space. It is convenient to use Slater determinants (made up from an ordered list of N of the $2M$ spin orbitals $\phi_1 \ldots \phi_{2M}$) for this purpose, as the set of all N-electron Slater determinants completely spans the antisymmetric many-particle space in which the solution must lie. It is common to choose the spin orbitals to be the Hartree–Fock (HF) orbitals, which lead to a natural excitation hierarchy from the HF determinant, the lowest energy determinant. Denoting the orbitals occupied in the HF determinant by indices $i, j, \ldots$ and the unoccupied (virtual) orbitals, $a, b, \ldots$, we may express the CI wave function as

$$\Psi_{\mathrm{CI}} = C_0 D_0 + \sum_{ia} C_i^a D_i^a + \sum_{\substack{i<j \\ a<b}} C_{ij}^{ab} D_{ij}^{ab} + \cdots, \quad (1)$$

where D_0 is the HF determinant and, e.g., D_{ij}^{ab} is the determinant where spin orbitals i and j are replaced by spin orbitals a and b. The coefficients C may be variationally determined by solving the set of projection equations

$$\langle D_0 | \hat{H} - E | \Psi_{\mathrm{CI}} \rangle = 0, \quad (2)$$

$$\langle D_i^a | \hat{H} - E | \Psi_{\mathrm{CI}} \rangle = 0, \quad (3)$$

$$\langle D_{ij}^{ab} | \hat{H} - E | \Psi_{\mathrm{CI}} \rangle = 0, \ldots, \quad (4)$$

to give C and E, the energy of Ψ_{CI}. This process is often performed iteratively, but in CIQMC it is performed stochastically by reexpressing the equations as a propagator. While the operator $(\hat{H} - E)$ projects the solution Ψ_{CI} to zero, the operator $1 - \tau(\hat{H} - E)$ (where τ is some positive real number) will project Ψ_{CI} to itself and has the same eigenfunctions. Furthermore, as seen in diffusion Monte Carlo calculations [6], with suitably small τ, applying this repeatedly to any wave function with a component of Ψ_{CI} will result in Ψ_{CI}. Letting $D_{\mathbf{j}}$ and $C_{\mathbf{j}}$ denote some generic determinants and coefficients, with this new operator we may reexpress (3) and (4), etc., as

$$C_{\mathbf{i}} - \tau \langle D_{\mathbf{i}} | \hat{H} - E | D_{\mathbf{i}} \rangle C_{\mathbf{i}} - \tau \sum_{\substack{\mathbf{j} \to \mathbf{i} \\ \mathbf{j} \neq \mathbf{i}}} \langle D_{\mathbf{i}} | \hat{H} | D_{\mathbf{j}} \rangle C_{\mathbf{j}} = C_{\mathbf{i}}, \quad (5)$$

where $\mathbf{j} \to \mathbf{i}$ requires that $D_{\mathbf{j}}$ be connected to $D_{\mathbf{i}}$ via the Hamiltonian (i.e., $H_{\mathbf{ij}} = \langle D_{\mathbf{i}} | \hat{H} | D_{\mathbf{j}} \rangle \neq 0$). We may consider

this process as two steps, and create a new set of coefficients from the old ones,

$$C_{\mathbf{i}} - \tau(H_{\mathbf{ii}} - E)C_{\mathbf{i}} \rightarrow C_{\mathbf{i}}, \tag{6}$$

$$C_{\mathbf{i}} - \tau \sum_{\substack{\mathbf{j}\rightarrow\mathbf{i} \\ \mathbf{j}\neq\mathbf{i}}} H_{\mathbf{ij}} C_{\mathbf{j}} \rightarrow C_{\mathbf{i}}. \tag{7}$$

To express this process stochastically, we may discretize the variables C by creating a population of "walkers," each of which carries a sign and which may be at any determinant. Over a number of steps, the mean signed population of walkers on determinant $\mathbf{j}$ will be proportional to $C_{\mathbf{j}}$. For the population of walkers to average to the CI wave function, their dynamics must obey (6) and (7). This is achieved through three processes.

(i) *Spawning*.—Equation (7) shows that if there is a walker at determinant $\mathbf{j}$ connected to $\mathbf{i}$ then, with a probability proportional to $\tau|H_{\mathbf{ij}}|$, it will spawn an appropriately signed walker at $\mathbf{i}$. To sample this more effectively, as all walkers are equivalent, we may perform this in reverse; iterating through the list of walkers, at each we may generate a random connected excitation with normalized probability $p_{\mathrm{gen}}(\mathbf{j}|\mathbf{i})$ and spawn a walker with probability $\tau|H_{\mathbf{ij}}|/p_{\mathrm{gen}}(\mathbf{j}|\mathbf{i})$ with sign $-\mathrm{sgn}(H_{\mathbf{ij}})$. The spawning may be considered to replace (7) by a step where each $D_{\mathbf{j}}$ spawns onto $D_{\mathbf{i}}$, written as $-\tau H_{\mathbf{ij}} C_{\mathbf{j}} + C_{\mathbf{i}} \rightarrow C_{\mathbf{i}}$.

(ii) *Birth and death*.—Equation (6) corresponds to the creation or destruction of a walker at the determinant. $p_d(\mathbf{i}) = -\tau(H_{\mathbf{ii}} - E)$ is a signed quantity indicating the probability of birth (positive) or death (negative) of a walker at determinant $\mathbf{i}$.

(iii) *Annihilation*.—If there is more than one walker on a determinant, they may be of opposite sign. Such pairs of oppositely signed walkers are removed.

The remaining variable E can be seen to have the effect of population control. If, at convergence, E is higher than the lowest eigenvalue, the population of walkers at each determinant will tend to increase, and, if lower, decrease. E can thus be altered dynamically to stabilize the population, and is termed the "shift." This algorithm has been seen to be extremely effective at recovering CI results in small systems, with the drawback that the number of walkers required for a stable population increases as a roughly constant fraction of the Hilbert space, which itself increases exponentially with basis size and number of electrons. This notwithstanding, it has produced results on larger systems than previously studied.

We now seek to answer the question, "Can stochastic methods be used to tackle coupled cluster theory in a similar manner?" In CC theory, instead of expressing the wave function as a sum of determinants and coefficients, it is expressed as an exponential, $\Psi_{\mathrm{CC}} = N_0 e^{\hat{T}} D_0$. N_0 determines the normalization and $\hat{T}$, the cluster operator, is a sum of coefficients and excitation operators,

$$\hat{T} = \sum_{i,a} t_i^a \hat{a}_i^a + \sum_{\substack{i<j \\ a<b}} t_{ij}^{ab} \hat{a}_{ij}^{ab} + \cdots, \tag{8}$$

where an excitation operator (which consists of a string of creation and annihilation operators) acting on a determinant gives a higher excited determinant, e.g., $\hat{a}_{jk}^{bc} D_i^a = D_{ijk}^{abc}$. Where no such excitation is possible, the result is zero. The application of single or combinations of excitation operators may require index reordering (and thus sign changes) to express the excitations in ordered form required of the determinants, e.g., $\hat{a}_i^b \hat{a}_j^a D_0 = D_{ij}^{ba} = -D_{ij}^{ab}$. Some care must be taken with respect to normalization, and we may choose to insert an arbitrary constant to our ansatz, $\Psi_{\mathrm{CC}} = N_0 e^{N_T \hat{T}} D_0$, with N_T determining the scaling of the cluster operator. Expanding this out gives

$$\Psi_{\mathrm{CC}} = N_0 D_0 + N_0 \sum_{ia} N_T t_i^a D_i^a + N_0 \sum_{\substack{i<j \\ a<b}} [N_T t_{ij}^{ab} + N_T^2 (t_i^a t_j^b - t_i^b t_j^a)] D_{ij}^{ab} + \cdots. \tag{9}$$

This is more conveniently written out in terms of an increasing number of excitations, $\mathbf{i}, \mathbf{j}, \ldots,$

$$\Psi_{\mathrm{CC}} = N_0 D_0 + N_0 \sum_{\mathbf{i}} N_T t_{\mathbf{i}} \hat{a}_{\mathbf{i}} D_0 + \frac{N_0}{2} \sum_{\mathbf{ij}} N_T^2 t_{\mathbf{i}} t_{\mathbf{j}} \hat{a}_{\mathbf{i}} \hat{a}_{\mathbf{j}} D_0 + \cdots. \tag{10}$$

It is convenient to set $N_T = 1/N_0$, which recovers a similar form to the CIQMC:

$$\Psi_{\mathrm{CC}} = N_0 D_0 + \sum_{\mathbf{i}} t_{\mathbf{i}} \hat{a}_{\mathbf{i}} D_0 + \frac{1}{2N_0} \sum_{\mathbf{ij}} t_{\mathbf{i}} t_{\mathbf{j}} \hat{a}_{\mathbf{i}} \hat{a}_{\mathbf{j}} D_0 + \cdots. \tag{11}$$

The CC equations very much resemble those of CI:

$$\langle D_0 | \hat{H} - E | \Psi_{\mathrm{CC}} \rangle = 0, \tag{12}$$

$$\langle D_i^a | \hat{H} - E | \Psi_{\mathrm{CC}} \rangle = 0, \tag{13}$$

$$\langle D_{ij}^{ab} | \hat{H} - E | \Psi_{\mathrm{CC}} \rangle = 0, \ldots, \tag{14}$$

and as before we expand these out with the new operator $1 - \tau(\hat{H} - E)$. As a simple example we consider a system with two occupied spin orbitals (i and j) and two virtual (a and b) spin orbitals. We denote $D_{\mathbf{n}} \equiv D_{ij}^{ab}$.

$$\langle D_{\mathbf{n}} | 1 - \tau(\hat{H} - E) | \Psi_{\mathrm{CC}} \rangle = \langle D_{\mathbf{n}} | \Psi_{\mathrm{CC}} \rangle \tag{15}$$

$$(1 - \tau(H_{\mathbf{nn}} - E))(t_i^a t_j^b - t_i^b t_j^a + t_{ij}^{ab}) - \tau \sum_{\substack{\mathbf{m}\rightarrow\mathbf{n} \\ \mathbf{m}\neq\mathbf{n}}} H_{\mathbf{nm}} \frac{\langle D_{\mathbf{m}} | \Psi_{\mathrm{CC}} \rangle}{N_0} = (t_i^a t_j^b - t_i^b t_j^a + t_{ij}^{ab}), \tag{16}$$

where we have divided through by the normalization in the last line. This expression is rather more complex than (5),

but shares the same form, and to discretize we sample the t amplitudes as populations of excitors. N_0 may be represented by a population on the Hartree–Fock determinant, and $t_{\mathbf{i}}$ by a population of discrete excitors representing excitation amplitudes. To formulate the dynamics, we must consider not single excitors (t amplitudes), but all possible products of them (denoted inside [ ]). In this case, the relevant products are $[t_i^a t_j^b]$, $[t_i^b t_j^a]$, and t_{ij}^{ab}, which are separate and in some cases composite ($[t_i^a t_j^b]$ and $[t_i^b t_j^a]$) dynamical objects. A key choice is that the dynamics of such clusters of excitors is expressed so as to only create *single* (noncomposite) excitors. This allows clusters to be generated stochastically on the fly, requiring only noncomposite excitors to be stored. The dynamics are achieved by sampling clusters of all possible sizes from the list of excitors and applying a simple set of spawning and death steps. For example, the following steps reproduce (16):

$$(1 - \tau(H_{\mathbf{nn}} - E))[t_i^a t_j^b] + t_{ij}^{ab} \rightarrow [t_i^a t_j^b] + t_{ij}^{ab}, \quad (17)$$

$$(1 - \tau(H_{\mathbf{nn}} - E))[-t_i^b t_j^a] + t_{ij}^{ab} \rightarrow [-t_i^b t_j^a] + t_{ij}^{ab}, \quad (18)$$

$$(1 - \tau(H_{\mathbf{nn}} - E))t_{ij}^{ab} \rightarrow t_{ij}^{ab}, \quad (19)$$

$$t_{ij}^{ab} - \tau \sum_{\substack{\mathbf{m}\rightarrow\mathbf{n} \\ \mathbf{m}\neq\mathbf{n}}} H_{\mathbf{nm}} \frac{\langle D_{\mathbf{m}}|\Psi_{\mathrm{CC}}\rangle}{N_0} \rightarrow t_{ij}^{ab}. \quad (20)$$

These dynamics modify the population of the single, not composite, excitors. For example, in (17), the "death" of $[t_i^a t_j^b]$ does not affect the populations of t_i^a and t_j^b, but has an effect by creating excitors of the "collapsed" version, t_{ij}^{ab}. In exactly the same way as CIQMC, spawning is performed in reverse. Looking at the case where $D_{\mathbf{m}} \equiv D_{kl}^{cd}$, (20) is replaced by

$$-\tau H_{\mathbf{nm}} t_{kl}^{cd} + t_{ij}^{ab} \rightarrow t_{ij}^{ab}, \quad (21)$$

$$-\tau H_{\mathbf{nm}}[t_k^c t_l^d] + t_{ij}^{ab} \rightarrow t_{ij}^{ab}, \quad (22)$$

$$-\tau H_{\mathbf{nm}}[-t_k^d t_l^c] + t_{ij}^{ab} \rightarrow t_{ij}^{ab}. \quad (23)$$

Again, clusters of excitors can only spawn onto single excitors. These rules result in the dynamics of single excitors (e.g., t_{ij}^{ab}) being the same as CIQMC. The clusters of excitors (e.g., $[t_i^a t_j^b]$) are sampled by selecting multiple excitors randomly and combining them. The resultant cluster is then collapsed into a single excitor which may die (i.e., create an excitor of opposite sign), and spawn. This collapse can be seen from (22) and (23) which effectively treat clusters of excitors as their collapsed version, and allow them to spawn only single excitors.

While we have expressed both the CI and CC in the full space of determinants and excitations, we may restrict this space to a subset, by restricting the level of excitation of $C_{\mathbf{i}}$ and $t_{\mathbf{i}}$. Fuller details of the implementation of this scheme will be given elsewhere, but we note a number of choices in the algorithm which differentiate it from the CIQMC. As the number of composites of excitors available is exponentially large, they must be sampled stochastically. Whereas in CIQMC, uniform sampling of walkers allows their generation probabilities to be ignored, care must be taken to produce normalized generation probabilities of the clusters of excitors. In CIQMC a projected energy, $E_{\mathrm{proj}} = \langle D_0|\hat{H}|\Psi\rangle/\langle D_0|\Psi\rangle$, is evaluated and used as an output or convergence test. In CC Monte Carlo (CCMC), there is a contribution to this from singles and doubles excitors as well as from products of singles excitors which is evaluated stochastically. Owing to the dependence of the cluster space on the population at the zero excitor, N_0, it is necessary to keep a relatively large population there, and so simulations are started with a population in the range 50–500, rather than from a single walker as in CIQMC.

We now study some small systems to investigate the efficacy of this new formulation. In particular, we wish to determine how the number of excitors affects the convergence of the energy, and if any systematic error is introduced by the discretization and sampling. We compare to CI calculations based on integrals from Q-CHEM [10], and CC calculations from the additions by Parkhill *et al.* [11]. Figure 1 shows how the projected energy of the water molecule converges with number of steps, achieving sub-milli-Hartree accuracy within a few thousand cycles.

In CIQMC there is no systematic error above a critical number of walkers, determined by a plateau in the walker growth. Below this number of walkers a systematic error is apparent and the two measures of energy (projected and the shift) do not agree. Table I shows plateau heights for

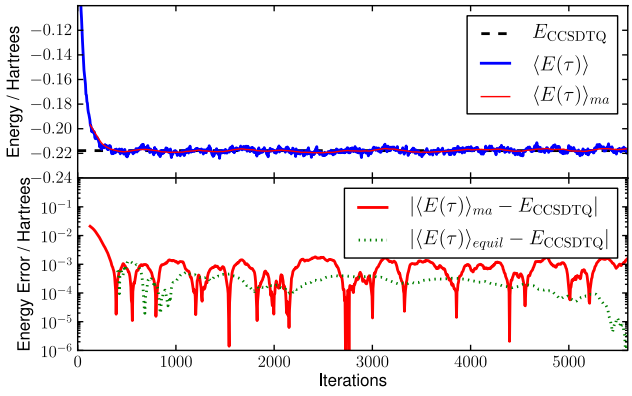

FIG. 1 (color online). CCMC for water in a cc-pVDZ basis at quadruples truncation. The simulation was started with a population of 4000 at the Hartree–Fock with $\delta\tau = 0.005$ a.u. $\langle\rangle_{\mathrm{ma}}$ indicates a moving average over 50 shift-steps each of $5\delta\tau$ and $\langle\rangle_{\mathrm{equil}}$ indicates a cumulative expectation value when the system has equilibrated after 400 steps.

TABLE I. CIQMC and CCMC correlation energies of various systems. Core electrons were frozen except for Ne (cc-pVDZ) and H_2O. The number of excitors was chosen to be above the level of the plateaux where found. MC energies are the mean of the projected energy. Errors were calculated using a blocking analysis [12]. CC data and geometries for H_2O are taken from Olsen *et al.* [13]. Calculations used $\tau = 0.001$ except for quadruples truncation for N_2 which had $\tau = 0.000\,01$.

System	Truncation level	# determinants	CIQMC plateau	CCMC plateau	CIQMC energy	CI energy	CCMC energy	CC energy
Ne cc-pVDZ	2	400		40	−0.1864(2)	−0.186 651	−0.1913(2)	−0.190 861
	3	4680		200	−0.1876(2)	−0.187 624	−0.1920(1)	−0.191 945
	4	30 654		700	−0.1922(2)	−0.191 989	−0.192 10(4)	−0.192 095
	5	113 550		3000	−0.192 09(4)	−0.192 071	−0.192 03(9)	−0.192 106
	10 (full)	501 992		6000	−0.192 13(4)	−0.192 106	−0.192 15(6)	−0.192 106
Ne aug-cc-pVDZ	2	985	70	200	−0.2050(3)	−0.205 120	−0.2102(3)	−0.210 154
	3	18 341	300	1700	−0.2076(2)	−0.207 625	−0.2131(2)	−0.213 066
	4	183 731	3900	50 000	−0.2129(2)	−0.212 918	−0.213 10(5)	−0.213 125
H_2O cc-pVDZ	2	3416	250	1200	−0.2057(2)	−0.205 798	−0.2140(3)	−0.214 077
	3	90 280	10 500	45 000	−0.2089(1)	−0.208 779	−0.2173(1)	−0.217 238
	4	1 291 578	160 000	600 000	−0.217 45(4)	−0.217 494	−0.217 80(8)	−0.217 802
N_2 cc-pVDZ eqm	2	1692	200	600	−0.2923(2)	−0.292 502	−0.3147(2)	−0.314 493
$r_{NN} = 2.118a_0$	3	44 340	9000	35 000	−0.3031(2)	−0.303 016	−0.3271(3)	−0.327 122
	4	634 980	173 000	392 000	−0.326 67(2)	−0.326 654	−0.328 65(11)	−0.328 732
N_2 cc-pVDZ str	2	1692	800	900	−0.3803(4)	−0.379 829	−0.4353(4)	−0.434 795
$r_{NN} = 3.0a_0$	3	44 340	12 500	25 000	−0.4023(3)	−0.402 326	−0.4694(5)	−0.469 673
	4	634 980	218 000	531 000	−0.465 58(5)	−0.465 565	−0.477 57(15)	−0.477 57

the systems studied. In common with CIQMC, CCMC calculations show a plateau in excitor growth at larger numbers of excitors than a CIQMC calculation of the same truncation. This is to be expected as the space of clusters of excitors in CCMC grows much more quickly than that of determinants in CIQMC. As the CCMC calculations are started with a significant population at the zero excitor to maintain stability, the plateau phase is often extremely short, and the nonexponential initial population growth reduces smoothly to the stable exponential growth expected of a calculation with the shift fixed at zero. We define the onset of this latter growth as the plateau height. Overall it can be seen that in the systems studied, the critical number of excitors at the plateau is in all cases smaller than the total size of the space, and owing to the relatively rapid convergence of the CC expansion with truncation level, many fewer excitors are required to recover energies of Full CI quality than the equivalent FCIQMC calculation.

In conclusion, we have formulated the coupled cluster equations in such a way as to sample them stochastically with discrete excitors. The population dynamics of these excitors recovers the appropriate CC energy within statistical error bars, requiring fewer excitors than amplitudes would be required to perform the exact calculation. Owing to the rapid developments of the CIQMC method, particularly the initiator modification [8], along with very parallelizable nature of MC methods, there are prospects for application of CC theory to far larger systems than previously possible, and we hope to perform these, as well as scaling studies, in the near future.

The author would like to thank Churchill College, Cambridge for support, and George Booth and Ali Alavi for helpful discussions and code.

*alex.thom@imperial.ac.uk

[1] K. Kowalski, D. J. Dean, M. Hjorth-Jensen, T. Papenbrock, and P. Piecuch, Phys. Rev. Lett. **92**, 132501 (2004).
[2] B. K. Sahoo, R. Chaudhuri, B. P. Das, and D. Mukherjee, Phys. Rev. Lett. **96**, 163003 (2006).
[3] D. J. J. Farnell, R. F. Bishop, and K. A. Gernoth, Phys. Rev. B **63**, 220402(R) (2001).
[4] R. J. Bartlett and M. Musiał, Rev. Mod. Phys. **79**, 291 (2007).
[5] A. J. W. Thom and A. Alavi, Phys. Rev. Lett. **99**, 143001 (2007).
[6] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[7] G. H. Booth, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **131**, 054106 (2009).
[8] D. Cleland, G. H. Booth, and A. Alavi, J. Chem. Phys. **132**, 041103 (2010).
[9] G. H. Booth and A. Alavi, J. Chem. Phys. **132**, 174104 (2010).
[10] Y. Shao *et al.*, Phys. Chem. Chem. Phys. **8**, 3172 (2006), Q-CHEM version 3.0.
[11] J. A. Parkill and M. Head-Gordon, Mol. Phys. **108**, 513 (2010).
[12] H. Flyvbjerg and H. G. Petersen, J. Chem. Phys. **91**, 461 (1989).
[13] J. Olsen, P. Jørgensen, H. Koch, A. Balkova, and R. J. Bartlett, J. Chem. Phys. **104**, 8007 (1996).