

Stochastic Coupled Cluster Theory

Alex J. W. Thom*

*Department of Chemistry, Imperial College London, London SW7 2AZ, U.K. and
University of Cambridge, Chemistry Department, Lensfield Road, Cambridge CB2 1EW, U.K.*

(Dated: December 4, 2010)

We describe a Stochastic Coupled Cluster Theory which represents excitation amplitudes as discrete excitors in the space of excitation amplitudes. Re-expressing the Coupled Cluster (CC) equations as the dynamics of ‘excitons’ in this space, we show that a simple set of rules suffice 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.

Coupled Cluster 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 favourable 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 ‘excitons’ 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 wavefunction 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 $2M$ spinorbitals $\phi_1 \dots \phi_{2M}$) 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 spinorbitals 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, \dots and the unoccupied (virtual) orbitals, a, b, \dots , we may express the CI wavefunction as

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

where D_0 is the HF determinant and, for example, D_{ij}^{ab} is the determinant where spinorbitals i and j have been replaced with spinorbitals a and b . The coefficients C may be variationally determined by solving the set of projection equations

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

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

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

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[6], with suitably small τ , applying this repeatedly to any wavefunction 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),(4), etc. as

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

where $\mathbf{j} \rightarrow \mathbf{i}$ 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,

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

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

To express this process stochastically, we may discretize the variables C by creating a population of ‘walkers’, each of which carries a sign and whose position may be at any determinant. Over a number of steps, the average 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 wavefunction, their dynamics must be such that they obey (6) and (7). This is achieved through three processes:

- (i) *Spawning*: (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 $p_{gen}(\mathbf{j}|\mathbf{i})$ and spawn a walker with probability $\tau|H_{\mathbf{ij}}|/p_{gen}(\mathbf{j}|\mathbf{i})$ with sign $-\text{sgn}(H_{\mathbf{ij}})$. 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*: (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*: When 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 wavefunction as a sum of determinants and coefficients, it is expressed as an exponential, $\Psi_{CC} = N_0 e^{\hat{T}} D_0$. N_0 determines the normalization of the wavefunction and \hat{T} , the cluster operator, is a sum of coefficients and excitation operators,

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

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. 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 owing to the form of the wavefunction. We may choose to insert an arbitrary constant to our *Ansatz*, $\Psi_{CC} = N_0 e^{N_T \hat{T}} D_0$, which allows more freedom in the discretization, with N_T determining the scaling of the cluster operator. Expanding this out gives

$$\begin{aligned} \Psi_{CC} = & N_0 D_0 + N_0 \sum_{ia} N_T t_i^a D_i^a \\ & + N_0 \sum_{\substack{i<j \\ a<b}} [N_T t_{ij}^{ab} + N_T^2 (t_i^a t_j^b - t_i^b t_j^a)] D_{ij}^{ab} + \dots \end{aligned} \quad (9)$$

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

$$\begin{aligned} \Psi_{CC} = & N_0 D_0 + N_0 \sum_{\mathbf{i}} N_T t_{\mathbf{i}} \hat{a}_{\mathbf{i}} D_{\mathbf{i}} \\ & + \frac{N_0}{2} \sum_{\mathbf{ij}} N_T^2 t_{\mathbf{i}} t_{\mathbf{j}} \hat{a}_{\mathbf{i}} \hat{a}_{\mathbf{j}} D_{\mathbf{i}\mathbf{j}} + \dots \end{aligned} \quad (10)$$

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

$$\begin{aligned} \Psi_{CC} = & N_0 D_0 + \sum_{\mathbf{i}} t_{\mathbf{i}} \hat{a}_{\mathbf{i}} D_{\mathbf{i}} \\ & + \frac{1}{2N_0} \sum_{\mathbf{ij}} t_{\mathbf{i}} t_{\mathbf{j}} \hat{a}_{\mathbf{i}} \hat{a}_{\mathbf{j}} D_{\mathbf{i}\mathbf{j}} + \dots \end{aligned} \quad (11)$$

The CC equations very much resemble those of CI:

$$\langle D_0 | \hat{H} - E | \Psi_{CC} \rangle = 0 \quad (12)$$

$$\langle D_i^a | \hat{H} - E | \Psi_{CC} \rangle = 0 \quad (13)$$

$$\langle D_{ij}^{ab} | \hat{H} - E | \Psi_{CC} \rangle = 0, \dots \quad (14)$$

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 spinorbitals (i and j) and two virtual (a and b) spinorbitals. We denote $D_{\mathbf{n}} \equiv D_{ij}^{ab}$.

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

$$(1 - \tau(H_{\mathbf{nn}} - E))(t_i^a t_j^b - t_i^b t_j^a + t_{ij}^{ab}) \quad (16)$$

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

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. To discretize this equation we discretize the t -amplitudes as populations of ‘excitons’. N_0 may be represented by a population on the Hartree–Fock determinant, and t_i by a population of discrete ‘excitons’ representing excitation amplitudes. To formulate the dynamics, we must consider not single excitons (t -amplitudes), but all possible products of them (denoted inside \square). 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 in this method is that the dynamics of such clusters of excitons is expressed so as to only create *single* (non-composite) excitons. This allows clusters to be generated stochastically on-the-fly, requiring only non-composite excitons to be stored. The dynamics are achieved by sampling clusters of all possible sizes from the list of excitons and applying a simple set of spawning and death steps. For example, the following steps reproduce (16):

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

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

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

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

These dynamics modify the population of the single, not composite, excitons. 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 excitons 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 excitons can only spawn onto single excitons. These rules result in the dynamics of single excitons (e.g. t_{ij}^{ab}) being the same as CIQMC. The clusters of excitons (e.g. $[t_i^a t_j^b]$) are sampled by selecting multiple excitons randomly and combining them. The resultant cluster is then collapsed into a single exciton which may die (i.e. create an exciton of opposite sign), and spawn. This ‘collapse’ can be seen from (22) and (23) which effectively treat clusters of excitons as their collapsed version, and allow them to spawn only single excitons.

We note that while we have expressed both the CI and CC in the full space of determinants and excitations, we have the liberty of restricting this space to a subset, by restricting the level of excitation of C_i and t_i . Fuller details of the implementation of this scheme will be given elsewhere, but we note a number of choices

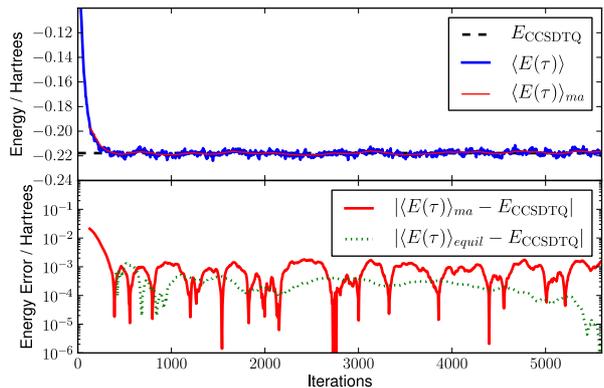


FIG. 1: CCMC for water in a cc-pVDZ basis at quadruples truncation. The simulation was started with a population of 4000 at the Hartree–Fock with $\delta\tau = 0.005$ a.u. $\langle \rangle_{ma}$ indicates a moving average over 50 shift-steps each of $5\delta\tau$ and $\langle \rangle_{equil}$ indicates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

in the algorithm which differentiate it from the CIQMC. As the number of composites of excitons 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 excitons. In CIQMC a projected energy, $E_{proj} = \langle D_0 | \hat{H} | \Psi \rangle / \langle D_0 | \Psi \rangle$, is evaluated and used as an output or convergence test. In CCMC, there is a contribution to this from singles and doubles excitons as well as from products of singles excitons. As the latter contribution be relatively expensive to compute, it is evaluated stochastically. Owing to the dependence of the cluster space on the population at the zero exciton, 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 excitons affects the convergence of the energy, as well as investigate if there is any systematic error introduced by the discretization and sampling. We compare to CI calculations based on integrals from Q-Chem[12], and CC calculations from the additions by Parkhill *et al.*[13]. Figure 1 shows how the projected energy of the water molecule converges with number of steps, achieving sub-milliHartree accuracy within a few thousand cycles.

In CIQMC it has been seen that above a critical number of walkers there is no systematic error, and that number can be determined by locating 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

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)	-0.186651	-0.1913(2)	-0.190861
	3	4680		200	-0.1876(2)	-0.187624	-0.1920(1)	-0.191945
	4	30654		700	-0.1922(2)	-0.191989	-0.19210(4)	-0.192095
	5	113550		3000	-0.19209(4)	-0.192071	-0.19203(9)	-0.192106
	10 (full)	501992		6000	-0.19213(4)	-0.192106	-0.19215(6)	-0.192106
Ne aug-cc-pVDZ	2	985	70	200	-0.2050(3)	-0.205120	-0.2102(3)	-0.210154
	3	18341	300	1700	-0.2076(2)	-0.207625	-0.2131(2)	-0.213066
	4	183731	3900	50000	-0.2129(2)	-0.212918	-0.21310(5)	-0.213125
H ₂ O cc-pVDZ	2	3416	250	1200	-0.2057(2)	-0.205798	-0.2140(3)	-0.214077
	3	90280	10500	45000	-0.2089(1)	-0.208779	-0.2173(1)	-0.217238
	4	1291578	160000	600000	-0.21745(4)	-0.217494	-0.21780(8)	-0.217802
N ₂ cc-pVDZ eqm $r_{NN} = 2.118a_0$	2	1692	200	600	-0.2923(2)	-0.292502	-0.3147(2)	-0.314493
	3	44340	9000	35000	-0.3031(2)	-0.303016	-0.3271(3)	-0.327122
	4	634980	173000	392000	-0.32667(2)	-0.326654	-0.32865(11)	-0.328732
N ₂ cc-pVDZ str $r_{NN} = 3.0a_0$	2	1692	800	900	-0.3803(4)	-0.379829	-0.4353(4)	-0.434795
	3	44340	12500	25000	-0.4023(3)	-0.402326	-0.4694(5)	-0.469673
	4	634980	218000	531000	-0.46558(5)	-0.465565	-0.47757(15)	-0.47757

TABLE I: CIQMC and CCMC correlation energies of various systems. Ne (cc-pVDZ) and H₂O correlated all electrons, and the remaining calculations had frozen cores. 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[10]. CC data and geometries for H₂O are taken from Olsen *et al.*[11]. Calculations used $\tau = 0.001$ except for quadruples truncation for N₂ which had $\tau = 0.00001$.

plateau heights for the systems studied. In common with CIQMC, CCMC calculations exhibit a plateau in their excitor growth, but often at much higher 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 must often be started with a significant population at the zero excitor to maintain stability, the plateau phase is often extremely short, and in effect the non-exponential 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 ‘exciters’. 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.

* Electronic address: 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. Mitras, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **75**, 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] H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.* **91**, 461 (1989).

- [11] J. Olsen, P. Jørgensen, H. Koch, A. Balkova, and R. J. Bartlett, *J. Chem. Phys.* **104**, 8007 (1996).
- [12] Y. Shao, M. Head-Gordon, et al., *Phys. Chem. Chem. Phys.* **8**, 3172 (2006), Q-Chem version 3.0.
- [13] J. A. Parkill and M. Head-Gordon, *Mol. Phys.* **108**, 513 (2010).