Exciting determinants in Quantum Monte Carlo: Loading the dice with fast, low memory weights

Verena A. Neufeld* and Alex J. W. Thom*

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

E-mail: van26@cam.ac.uk; ajwt3@cam.ac.uk

Abstract

High-quality excitation generators are crucial to the effectiveness of coupled cluster Monte Carlo (CCMC) and full configuration interaction Quantum Monte Carlo (FCIQMC) calculations. The heat bath sampling of Holmes et al. [Holmes, A. A.; Changlani, H. J.; Umrigar, C. J. *J. Chem. Theory Comput.* **2016**, *12*, 1561–1571.] dramatically increases the efficiency of the spawn step of such algorithms but requires memory storage scaling quartically with system size which can be prohibitive for large systems. Alternatively, Alavi *et al.* [Smart, S. D.; Booth, G. H.; Alavi, A., unpublished] approximated these weights with weights based on Cauchy–Schwarz-like inequalities calculated on-the-fly. While reducing the memory cost, this algorithm scales linearly in system size computationally. We combine both of these ideas with the single reference nature of many systems studied, and introduce a spawn-sampling algorithm that has low memory requirements (quadratic in basis set size) compared to the heat bath algorithm and only scales either independently of system size (CCMC) or linearly in the number of electrons (FCIQMC) that works especially well on localized orbitals. Tests on small water chains with localized orbitals with CCMC and with an initiator point sample in FCIQMC indicate that it can be equally efficient as the other excitation generators. As the system gets larger, calculations with our new algorithm converge faster

than the on-the-fly weight algorithm, while having a much more favourable memory scaling than the heat bath algorithm.

1 Introduction

Coupled cluster Theory[1–4] can give ground state energies to chemical accuracy (1 kcal mol^{-1})[4,5] in a systematically improvable manner. As an alternative to deterministically solving the coupled cluster equations, stochastic coupled cluster (CCMC)[6–10] adopts a sparse representation of the wavefunction which can reduce memory requirements compared to a full deterministic representation. This enables the use of higher coupled cluster levels and larger basis sets. Recently,[11] a finite uniform electron gas has been studied with truncation levels coupled cluster singles and doubles (CCSD) up to quintuple excitations (CCSDTQ5) reaching basis set sizes of 18342 spinorbitals. Like full configuration interaction quantum Monte Carlo (FCIQMC),[12,13] CCMC also does not suffer from the fermion sign problem in the same way[14] as diffusion Monte Carlo (DMC)[15] does. Provided there are enough walkers/Monte Carlo particles in the calculations,[13,14] FCIQMC energies are exact for the basis set chosen. FCIQMC has been applied to various molecules[16–23] and some periodic systems[11,14,24–30] to find ground state energies. It has also been used to determine excited state energies for example.[31–35] Both CCMC and

FCIQMC have been used with the CC(P;Q) technique,[36] which can speed up the time needed to find the main excitors/determinants in CC(P;Q). The algorithm used to perform FCIQMC and CCMC affects the speed and convergence, and there is still great scope for improvements.[7–9,13,20,37–41] Here, we propose a change to the *spawn* step in the algorithm to use weighted excitations, inspired by the heat bath algorithm proposed by Holmes et al.[20] (which was then expanded to the heat bath configuration interaction algorithm[42–44]), and Cauchy–Schwarz weights proposed by Smart et al.[40] The method introduced here has a lower computational scaling in CCMC than the heat bath excitation generators and a significantly lower memory cost.

The main part of Fock space quantum Monte Carlo algorithms such as CCMC and FCIQMC consists of four steps; *selection* of a determinant/an excitor, *spawn*, *death* and *annihilation*. The *spawn* part of the algorithm explores the space of possible determinants/excitors. For a given determinant, it decides how the determinants connected via the hamiltonian become involved in the wavefunction by becoming occupied. As Holmes et al.[20] already noted, it is not efficient to give all connected determinants/excitors an equal probability of being considered as some are more important for the dynamics than others. They have shown that their heat bath weighting when selecting states to spawn to can greatly improve the overall efficiency.

A Slater determinant $|D_{\mathbf{m}}\rangle$ is connected to another determinant $|D_{\mathbf{n}}\rangle$ by their connecting Hamiltonian element $\langle D_{\mathbf{n}}| \hat{H} |D_{\mathbf{m}}\rangle$ as part of the *spawn* step and the algorithms to choose $|D_{\mathbf{n}}\rangle$ given $|D_{\mathbf{m}}\rangle$ are called excitation generators. The probability of this generation is denoted $p(\mathbf{n}|\mathbf{m}) = p_{\text{gen}}$ after which a spawn occurs with probability $p_{\text{spawn}} \propto \delta\tau \frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$, with time step $\delta\tau$.

For an efficient calculation, p_{spawn} should have a reasonable value. If $p_{\text{spawn}} > 1$, multiple particles are spawned at the same time, known as a "bloom", which is undesirable as it leads to less controllable population dynamics. If, however, p_{spawn} is small, determinants are selected with no resulting spawn, and the algorithm is inefficient. p_{spawn} therefore ideally has a constant value, which can be altered by the time step $\delta\tau$. Hence, it is desirable to make p_{gen} proportional to $|\langle D_{\mathbf{n}}| \hat{H} |D_{\mathbf{m}}\rangle|$ rather than selecting determinants uniformly. Holmes et al.[20] have introduced a heat bath sampling algorithm which weights the probability of choosing $|D_{\mathbf{n}}\rangle$ with approximately $|\langle D_{\mathbf{n}}| \hat{H} |D_{\mathbf{m}}\rangle|$, but requires precomputation of Hamiltonian elements leading to a significant storage cost of $\mathcal{O}(M^4)$ (which is of the same order as stored integrals if the code does not calculate them on-the-fly) and computational cost of $\mathcal{O}(N)$ where M and N are the size of the basis set and number of electrons respectively. Smart et al.[40] have reported the use of the Cauchy–Schwarz-like inequalities to provide upper bounds for $|\langle D_{\mathbf{n}}| \hat{H} |D_{\mathbf{m}}\rangle|$ with weights calculated on-the-fly. This reduces the storage cost while being linearly scaling in the number of orbitals.

Inspired by these ideas, excitation generators were investigated with weights generated on-the-fly using Cauchy–Schwarz and Power–Pitzer[45] inequalities to approximate $|\langle D_{\mathbf{n}}| \hat{H} |D_{\mathbf{m}}\rangle|$ whose computational cost scales linearly with the number of spinorbitals in the basis, M. We then present a new excitation generator that uses this Power–Pitzer inequality but is of low computational order, $\mathcal{O}(N_{\text{ex.}})$ in the case of CCMC or $\mathcal{O}(N)$ when using FCIQMC, with memory cost $\mathcal{O}(M^2)$ which is also below the heat bath memory scaling. $N_{\text{ex.}}$ for a determinant or excitor is the number of electrons excited with respect to the reference. For a truncated coupled cluster theory $N_{\text{ex.}}$ does not scale with system size. In a single-reference calculation, the reference determinant carries the most weight in the wavefunction and the majority of spawnings occur from determinants within a few electrons of excitation of this. We therefore may pre-compute excitation weightings based on the reference determinant, which shares the majority of electrons with nearby excited determinants, and then map the excitation to apply to any excited determinant, $|D_{\mathbf{n}}\rangle$. By this method, similar weights to the heat bath algorithm and weights inspired by a

Power–Pitzer inequality are employed and the spread in $\frac{|\langle D_\mathbf{n}|\hat{H}|D_\mathbf{m}\rangle|}{p_\mathrm{gen}}$ is minimised at a reduced computational and memory cost.

We now give a brief summary of the CCMC method, followed by a more detailed description of various excitation generators whose performances we then compare.

2 Coupled Cluster Monte Carlo

This section describes the coupled cluster Monte Carlo (CCMC) method. Full configuration interaction Quantum Monte Carlo (FCIQMC) has been discussed extensively in the literature, see e.g. Refs.[12,13,27] Note that CCMC and FCIQMC can both use the same excitation generators. CCMC solves the coupled cluster equations stochastically enabling calculations with larger basis sets and coupled cluster truncation levels than deterministic methods. This section gives an overview over the algorithm and more information can be found in Refs.[6–9]

The single reference coupled cluster wavefunction $|\Psi\rangle$ is written as

$$|\Psi\rangle \propto \exp\left(\hat{T}\right)|D_\mathbf{0}\rangle \qquad (1)$$

where $|D_\mathbf{0}\rangle$ is the reference determinant and

$$\hat{T} = \sum_\mathbf{i} t_\mathbf{i}\hat{a}_\mathbf{i}. \qquad (2)$$

$\hat{a}_\mathbf{i}$ are "excitors" that generate determinants from the reference as

$$|D_\mathbf{i}\rangle = \hat{a}_\mathbf{i}|D_\mathbf{0}\rangle. \qquad (3)$$

The one-electron orbitals used here are all orthogonal. The set of $\hat{a}_\mathbf{i}$ included depends on the coupled cluster truncation level. In coupled cluster singles and doubles (CCSD), those that create single and double excitations are included whereas CCSDT contains those with triple excitations as well and so on. The unconventional unlinked coupled cluster equations are solved as,[46]

$$\langle D_\mathbf{n}|\,\hat{H}\,|\Psi\rangle = E\,\langle D_\mathbf{n}|\Psi\rangle \qquad (4)$$

for the ground state energy E. Multiplying by a small number, $\delta\tau$, this can be rewritten as

$$\langle D_\mathbf{n}|\,1 - \delta\tau(\hat{H} - E)\,|\Psi\rangle = \langle D_\mathbf{n}|\Psi\rangle. \qquad (5)$$

with imaginary time step $\delta\tau$. We arrive at an iterative equation for the coefficients $t_\mathbf{i}$ for the Slater determinants in Ψ in imaginary time τ

$$t_\mathbf{n}(\tau + \delta\tau) = t_\mathbf{n}(\tau) - \delta\tau\,\langle D_\mathbf{n}|\,(\hat{H} - E)\,|\Psi(\tau)\rangle. \qquad (6)$$

Franklin et al.[8] have rewritten equation 6 as

$$\begin{aligned} t_\mathbf{n}(\tau + \delta\tau) =&\, t_\mathbf{n}(\tau) \\ &- \delta\tau\,\langle D_\mathbf{n}|\,(\hat{H} - E_\mathrm{proj.} - E_\mathrm{HF})\,|\Psi(\tau)\rangle \\ &- \delta\tau(E_\mathrm{proj.} - S)t_\mathbf{n}(\tau). \end{aligned} \qquad (7)$$

where the sum of the Hartree–Fock energy E_HF and the shift S was substituted for the ground state energy E. The projected energy $E_\mathrm{proj.}$ and the population controlling shift S (described below) are both measures for the correlation energy. $E_\mathrm{proj.}$ is given by

$$E_\mathrm{proj.} = \frac{\langle D_\mathbf{0}|\,\hat{H} - E_\mathrm{HF}\,|\Psi\rangle}{\langle D_\mathbf{0}|\Psi\rangle}. \qquad (8)$$

Equation 7 is then sampled stochastically as described below and $t_\mathbf{n}$ is updated at each time step. Monte Carlo particles, "excips", are placed on the excitors $a_\mathbf{i}$. At first, all excips are on the null excitor $a_\mathbf{0}$ that gives back the reference determinant. As the simulation proceeds, they multiply and spread to other excitors with *spawn*, *death/birth* and *annihilation* steps at each imaginary time step.

During each time step, a single excitor or cluster of excitors which have particles on them are first randomly selected, e.g. the two excitors $a_\mathbf{i}$ and $a_\mathbf{j}$, that when acting together on the reference determinant $|D_\mathbf{0}\rangle$, give another determinant $|D_\mathbf{m}\rangle$, i.e. $\hat{a}_\mathbf{i}\hat{a}_\mathbf{j}|D_\mathbf{0}\rangle = \hat{a}_\mathbf{i}|D_\mathbf{j}\rangle = |D_\mathbf{m}\rangle$. Note that in the full non-composite algorithm,[10] this is selected slightly differently. This determinant then undergoes the following processes:

- *Spawn:* Another determinant $|D_\mathbf{n}\rangle$ is randomly selected with a probability p_gen. An excip of appropriate sign is placed on $\hat{a}_\mathbf{n}$ with a probability proportional to $\frac{\delta\tau|\langle D_\mathbf{n}|\hat{H}|D_\mathbf{m}\rangle|}{p_\text{gen}}$.

- *Death/Birth:* An excip of opposite or the same sign is placed on $a_\mathbf{m}$ with a probability proportional to $|\langle D_\mathbf{m}|\hat{H} - S - E_\text{HF}|D_\mathbf{m}\rangle|$ if just one excitor was used to form $|D_\mathbf{m}\rangle$ and a probability proportional to $|\langle D_\mathbf{m}|\hat{H} - E_\text{proj.} - E_\text{HF}|D_\mathbf{m}\rangle|$ if a cluster was used.

- *Annihilation:* Finally, at the end of a imaginary time step, excip pairs of opposite sign on the same excitor are removed.

The shift is initially set to zero and is allowed to vary once the total population (number of particles), $N_\text{tot.}$, is higher than the critical population at the "shoulder" or "plateau".[6,14] To give an on-average constant population, it is updated every B iterations according to

$$S(\tau) = S(\tau - \delta\tau B) - \frac{\gamma}{B\delta\tau} \ln \frac{N(\tau)}{N(\tau - \delta\tau B)} \quad (9)$$

where γ is the shift damping factor.

Rather than integer-valued, real-valued excip amplitudes[37,47] have been used and the full non-composite version of the CCMC algorithm[10] with truncated and even selection[9] has been applied. We have varied the shift damping automatically to reduce the variance of the projected energy[1]. We have also used MPI and OpenMP parallelization.[10] The results here were checked for population control bias using a reweighting scheme by Umrigar et al.[48] and Vigor et al.[49] Data has been reblocked[50] implemented in pyblock[2] to estimate error bars. Our CCMC and FCIQMC calculations were done with the HANDE code[51] which is open source[3]. The integral files needed were

---

[1] This feature has been implemented by Charles Scott.

[2] For code, see https://github.com/jsspencer/pyblock

[3] See http://www.hande.org.uk/ and https://github.com/hande-qmc/hande for information and code

created with PySCF.[52] When applicable, localization has been applied using a Boys[53] localization function in PySCF.[52] Figures have been drawn with Matplotlib[54] and/or Inkscape[4].

3 Excitation Generators

As mentioned above, in the *spawn* step, the excitation generator selects a determinant $|D_\mathbf{n}\rangle$ connected to $|D_\mathbf{m}\rangle$ with probability p_gen. The spawn probability is proportional to $\frac{\delta\tau|\langle D_\mathbf{n}|\hat{H}|D_\mathbf{m}\rangle|}{p_\text{gen}}$. In this paper, we present a method that aims to use an optimal p_gen so that more important determinants are selected with a higher probability. An introduction to excitation generators in FCIQMC is given by Booth et al.[12,19] The idea of excitation generation and dividing by the generation probability was also discussed in e.g. Refs.[20,55–58] and a transition with uniform selection is also done by the configuration state function projector Monte Carlo method of Ohtsuka et al.[59] Kolodrubetz et al.[57] used a weighted excitation generator that — among other distributions — used the inverse momentum squared as a weight. Booth et al.[19] also considered weighting the excitation generation by Hamiltonian matrix elements by enumerating a subset of excitations with the magnitudes of these Hamiltonian elements. Due to the cost of finding $p_\text{gen.}$, this idea was not pursued further. A version of the uniform excitation generators described here, is explained in detail in Ref.[19]

The spawn probability is only non-zero if $\langle D_\mathbf{n}|\hat{H}|D_\mathbf{m}\rangle$ is non-zero. The Hamiltonians, $\hat{H}$, considered here only contain constant, one body, and two body terms. $\langle D_\mathbf{n}|\hat{H}|D_\mathbf{m}\rangle$ can therefore only be non-zero if $|D_\mathbf{n}\rangle$ and $|D_\mathbf{m}\rangle$ differ by at most two orbitals. To select a suitable $|D_\mathbf{n}\rangle$ for $|D_\mathbf{m}\rangle$ to spawn to, we can create a single or a double excitation from $|D_\mathbf{m}\rangle$ to generate $|D_\mathbf{n}\rangle$ ($\mathbf{n} \neq \mathbf{m}$). Any other excitation would lead to a zero spawn probability. Except for the "original" heat bath excitation generator, all excitation generators discussed here create a single or double excitation from $|D_\mathbf{m}\rangle$ to gen-

---

[4] https://inkscape.org/

erate $|D_\mathbf{n}\rangle$ with probability p_{single} or $1 - p_{\text{single}}$ respectively. As suggested by Holmes et al.,[20] we aim to appropriately select $p_{\text{spawn,single}}$ and $p_{\text{spawn,double}}$ by setting p_{single} suitably to optimize the distribution of excitations. For a single excitation where electron in spinorbital i is excited to spinorbital a,

$$p_{\text{gen,single}} = p_{\text{single}} p_{\text{method}} p(i) p(a|i). \qquad (10)$$

p_{method} contains additional factors depending on the selection method of i and a.

In the case of a double excitation, $ij \rightarrow ab$, as i and j ideally come from the same set of orbitals (those occupied in the determinant) and so do a and b (those unoccupied in the determinant), first ij and then ab are selected in all excitation generators discussed here. That means that for example while the selection order between i and j can vary, a will not be selected before either i and j. The possible orders are therefore $ijab$, $ijba$, $jiab$ and $jiba$. While the first selected occupied is called i and the second j, their indistinguishability has to be taken into account when calculating p_{gen}:

$$p_{\text{gen,double}} =$$
$$(1 - p_{\text{single}}) p_{\text{method}} (p(i)p(j|i)p(a|i,j)p(b|a,i,j)+$$
$$p(i)p(j|i)p(b|i,j)p(a|b,i,j)+$$
$$p(j)p(i|j)p(a|j,i)p(b|a,j,i)+$$
$$p(j)p(i|j)p(b|j,i)p(a|b,j,i)).$$
$$(11)$$

In a rather basic implementation, the spinorbitals with electrons to excite i (and j) and the spinorbitals to excite to a (and b) are selected with uniform probabilities. The excitation generator that we call *not renormalised excitation generator* or simply *no. renorm.* here, when doing a double excitation, first selects i and j as a pair with uniform probability from the set of occupied orbitals. In that case,

$$p_{\text{method}}(p(i)p(j|i) + p(j)p(i|j)) = \frac{2}{N(N-1)}, \qquad (12)$$

where the number of electron is N. If both i and j have the same spin, σ, then a is uniformly chosen from the set of virtual orbitals of that spin, otherwise it can be any virtual orbital. b is then selected uniformly from the set of orbitals (excluding a) with required spin and symmetry. With this selection of b, it is possible that after the selection of i, j, and a, there are no possible selections of b, or that in fact an occupied orbitals has been selected as b. In such cases, it is a forbidden excitation generation. In that case the spawn attempt will be unsuccessful (we set $|\langle D_\mathbf{m}| H |D_\mathbf{n}\rangle| = 0$).

The choice of how to select which electrons to excite and to which spinorbitals they are excited is is entirely arbitrary (assuming all valid excitations are possible), as long as the probability with which this selection has been done is known and p_{gen} is then calculated accordingly. As an alternative to the *not renormalised excitation generator (no. renorm.)*, most forbidden excitations (which lead to unsuccessful spawns) can be avoided by generating a different excitation and renormalising the appropriate probabilities. This is called the *renormalised excitation generator* or in short, *renorm.*. Again, see Booth et al.[12,19] for an in-depth description of uniform excitation generators.

In the following subsections, we describe the heat bath excitation generators and the *heat bath/uniform Power–Pitzer* excitation generators which follow the ideas of Alavi and others. Finally, the *heat bath Power–Pitzer ref.* excitation generator is presented, which pre-computes some weights based on the reference determinant and therefore has a very low computational cost not scaling with system size ($\mathcal{O}(N_{\text{ex.}})$ when using CCMC or scaling as $\mathcal{O}(N)$ for FCIQMC instead of $\mathcal{O}(M)$). Its memory cost is significantly less than *heat bath* excitation generators, being $\mathcal{O}(M^2)$ instead of $\mathcal{O}(M^4)$. In appendix A, further uniform excitation generators are discussed.

Table 1 gives an overview over the weighted excitation generators presented here. This table should be understood together with the following descriptions in the next subsections.

Table 1: Overview of weighted excitation generators. C.–S. means Cauchy–Schwarz and P.–P. Power–Pitzer. p.c. is pre-calculated and o.t.f. means on-the-fly. Comp./mem. $\mathcal{O}$ is the computational/memory order the excitation generator scales with. As a method of selection, "heat bath" refers to "selecting those like the heat bath excitation generator". Single excitations or ij in a double excitation that have been selected "uniformly", have been selected with the uniform *renorm.* excitation generator. N is number of electrons, M the number of spinorbitals and $N_{\text{ex.}}$ the excitation level possible from the reference at this coupled cluster level.

	single excitations	ij	ab	comp. $\mathcal{O}$	mem. $\mathcal{O}$
heat bath	decision after having selected ija	heat bath	heat bath	N	M^4
heat bath uniform singles	uniformly	heat bath	heat bath	N	M^4
heat bath exact singles	exactly, on-the-fly	heat bath	heat bath	N^2M or $NN_{\text{ex.}}M$	M^4
uniform C.–S.	uniformly	uniformly	C.–S. o.t.f.	M	M
uniform P.–P.	uniformly	uniformly	P.–P. o.t.f.	M	M
heat bath C.–S.	uniformly	heat bath	C.–S. o.t.f.	M	M^2
heat bath P.–P.	uniformly	heat bath	P.–P. o.t.f.	M	M^2
heat bath P.–P. ref.	p.c.	heat bath p.c.	P.–P. p.c.	N or $N_{\text{ex.}}$	M^2

3.1 Heat Bath Excitation Generators

The heat bath excitation generators aim to get the orbital selection weights as close as possible to the Hamiltonian matrix element $|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|$ with the aim of making part of the spawn probability $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$ as close as possible to a constant. In the case of a double excitation $ij \to ab$, p_{gen} can be rewritten as

$$p_{\text{gen,double}} = p(i) \times p(j|i) \times p(a|ij) \times p(b|ija) =$$
$$\frac{\sum_{jab} H_{ijab}}{\sum_{ijab} H_{ijab}} \times \frac{\sum_{ab} H_{ijab}}{\sum_{jab} H_{ijab}} \times \frac{\sum_{b} H_{ijab}}{\sum_{ab} H_{ijab}} \times \frac{H_{ijab}}{\sum_{b} H_{ijab}}, \quad (13)$$

where $H_{ijab} = |\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|$ where $|D_{\mathbf{m}}\rangle$ and $|D_{\mathbf{n}}\rangle$ differ by the excitation $ij \to ab$. In the case of heat bath excitation generators, $\frac{\sum_{jab} H_{ijab}}{\sum_{ijab} H_{ijab}}$ with certain limits in the sums is an approximation for $p(i)$ and so on.

Here, we distinguish between three different heat bath excitation generators described by/based on Holmes et al.[20] The "original" heat bath excitation generator as introduced and described in detail by Holmes et al.[20] (in short *heat bath*), the heat bath excitation generator that decides first whether a single or a double excitation is performed and which samples singles uniformly which is mentioned by Holmes et al.[20] (*heat bath uniform singles*) and finally, the one that first decides whether to do a single or double excitation and samples singles exactly according to their Hamiltonian matrix element, *heat bath exact singles*[5]. For more information and an in-depth description, see Ref.[20]

In all three heat bath excitation generators, all possible contractions of H_{ijab} appearing in equation 13 are pre-computed and stored. More specifically, $H_i = \sum_{jab} H_{ijab}$, $H_{ij} = \sum_{ab} H_{ijab}$, $H_{ija} = \sum_{b} H_{ijab}$ and H_{ijab} are pre-computed where i, j, a and b can be any spinorbital in the calculation. In all sums $i \neq j \neq a \neq b$. The alias method[20,60–63] is used and alias tables are pre-calculated for selecting a (given ij) with weights H_{ija} and one for selecting b (given ija) with weights H_{ijab} (which is of $\mathcal{O}(M^4)$). The look-up time with the alias method is of $\mathcal{O}(1)$. The alias tables for selecting i and selecting i given j are computed on-the-fly using pre-computed weights in $\mathcal{O}(N)$ time. The alias table for selecting i then only considers H_i from the set of occupied orbitals for i and when se-

---

[5]Idea by Alavi and co–workers, this was suggested to us as an alternative by Pablo López Ríos (personal communication).

lecting j given i, the alias table only considers H_{ij} with occupied j.

When using the *heat bath* excitation generator to find an excitation, first an alias table is created on-the-fly for i as described above and then i is selected. We proceed similarly for j. Using the pre-computed alias table with weights H_{ija}, a is found. If this orbital is occupied, we have a forbidden excitation and the spawn attempt was unsuccessful. Only at this stage it is decided whether to attempt a single or a double excitation. In the algorithm by Holmes et al.,[20] a single excitation is attempted with probability $\frac{H_{ia}}{H_{ia}+H_{ija}}$ and a double excitation is attempted with probability $\frac{H_{ija}}{H_{ia}+H_{ija}}$ if $H_{ia} < H_{ija}$ where $H_{ia} = |\langle D_{\mathbf{k}}| \hat{H} |D_{\mathbf{m}}\rangle|$ with $|D_{\mathbf{m}}\rangle$ and $|D_{\mathbf{k}}\rangle$ connected by the excitation $i \to a$. However, if $H_{ia} > H_{ija}$, both a double and a single excitation are attempted[6]. In our implementation in HANDE,[51] that approach was modified to only allow one excitation attempt per excitation generator call. If $H_{ia} \geq H_{ija}$, instead of choosing to attempt a single ($i \to a$) and a double ($ij \to ab$) excitation, a single or a double excitation is attempted with probability $\frac{1}{2}$ respectively. The rest follows Holmes et al.[20] Either a single excitation $i \to a$ is attempted now or b is selected from pre-computed weights and a double excitation $ij \to ab$ (provided b is not occupied) is attempted.

The *heat bath* excitation generator relies on single excitations being less significant. It has the major drawback in that it potentially has a bias if there exists no j to be selected after i and before a if $i \to a$ is valid. This is explained in more detail in Ref.[20] Our conservative but robust test for bias as implemented in HANDE, counts the number of j for which $\sum_b H_{ijab}$ is non zero for given ia. If this number is greater than the number of virtual orbitals, then there will always be an occupied j to be selected for allowed $i \to a$ and there is no bias. :q

---

[6]It is not clear from Holmes et al.[20] what happens if $H_{ia} = H_{ija}$

3.2 On-the-fly Power–Pitzer Excitation Generators

While bringing $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$ closer to a constant as uniform excitation generators,[20] *heat bath* excitation generators suffer from a large memory cost ($\mathcal{O}(M^4)$). Alavi and Smart et al.[40] had the idea of calculating approximate weights on-the-fly in $\mathcal{O}(M)$ calculation time which has a lower memory cost. This is for example mentioned by Blunt et al.,[35] Holmes et al.[20] and Schwarz.[64] They proposed calculating Cauchy–Schwarz-like upper bounds on the two body part of the Hamiltonian on-the-fly when doing a double excitation. Here, we also describe an excitation generator that uses an inequality derived by Power and Pitzer[45] instead. It effectively differs from *Cauchy–Schwarz* excitation generators as described here by the usage of exchange rather than Coulomb integrals. There is little published documentation about existing *Cauchy–Schwarz* excitation generators. A brief description of a first generation of these excitation generators is given by Schwarz,[64] and these have also been reported by Alavi et al.[7] though a full description of the precise algorithm has yet to be published so our implementation will almost certainly differ from existing versions.

Given that i, j, a and b are different, the only non-zero part of the Hamiltonian element $\langle D_{\mathbf{m}}| H |D_{\mathbf{n}}\rangle$ in a double excitation are the Coulomb integral $\langle ij|ab\rangle$ and the exchange integral $\langle ij|ba\rangle$ according to Slater-Condon rules.[65,66] Here, the notation

$$\langle ij|ab\rangle = \int \frac{\phi_i^*(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2)\phi_a(\mathbf{r}_1)\phi_b(\mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|}, \tag{14}$$

is used with one electron orbitals/spinorbitals ϕ that make up Slater determinants $|D_{\mathbf{x}}\rangle$. en example of such a weight used by Alavi and others for $ij \to ab$ is a Cauchy–Schwarz upper bound on $\langle ij|ab\rangle$ given by

$$\sqrt{|\langle ia|ia\rangle| \, |\langle jb|jb\rangle|} \geq |\langle ij|ab\rangle|. \tag{15}$$

The weights are such that a can be chosen (al-

---

[7]Personal communication with Ali Alavi and Pablo López Ríos.

most) independently of b and vice versa which makes the algorithm linear scaling in the number of spinorbitals. A Power–Pitzer[45] inequality is

$$\sqrt{|\langle ia|ai\rangle|\,|\langle jb|bj\rangle|} \geq |\langle ij|ab\rangle|. \qquad (16)$$

Exchange integrals are lower or equal in magnitude than Coulomb integrals (see e.g. Ref.[67]) which means that exchange integrals are the tighter upper bound for $|\langle ij|ab\rangle|$. The two body term in the Hamiltonian is $\langle ij|ab\rangle - \langle ij|ba\rangle$. When a and b have opposite spin, the two body term reduces to $\langle ij|ab\rangle$ and its Power–Pitzer upper bound is used as the weight. If a and b have the same spin, both orderings, ab and ba will generate the same excitation, and this is included in p_{gen}. This section gives a detailed description of the algorithm.

i and j can be selected uniformly or with the *heat bath* weightings producing a family of excitation generators. We denote by *uniform Cauchy–Schwarz* and *uniform Power–Pitzer* excitation generators which select them uniformly, like the *renorm.* excitation generator, and by *heat bath Cauchy–Schwarz* and *heat bath Power–Pitzer* those which select them as the *heat bath* excitation generators do with pre-calculated weights with memory cost of $\mathcal{O}(M^2)$[8]. The computational scaling is $\mathcal{O}(M)$ in both cases.

The *Power–Pitzer* and *Cauchy–Schwarz* excitation generators first decide whether to attempt a single or a double excitation according to p_{single}. For single excitations, the *renorm.* excitation generator is employed. When attempting double excitations, i and j are selected either uniformly or with *heat bath* weights out of the occupied orbitals of $|D_\mathbf{m}\rangle$. Then, a is selected out of the set of virtual spinorbitals $a_{\sigma_i,\mathrm{virt.}}$ with the same spin as i. a is selected with the probability of

$$p(a|ij) = p(a|i) = \frac{\sqrt{|\langle ia|ai\rangle|}}{\sum_{a=a_{\sigma_i,\mathrm{virt.}}} \sqrt{|\langle ia|ai\rangle|}} \qquad (17)$$

----

when using *Power–Pitzer* excitation generators or

$$p(a|ij) = p(a|i) = \frac{\sqrt{|\langle ia|ia\rangle|}}{\sum_{a=a_{\sigma_i,\mathrm{virt.}}} \sqrt{|\langle ia|ia\rangle|}} \qquad (18)$$

when using *Cauchy–Schwarz* excitation generators. b, the second orbital to excite to, it selected out of the set of spinorbitals $b_{\neq a,\sigma_j,\mathrm{sym.}}$ of the same spin as j and the required symmetry to conserve overall symmetry and not equal to a. The weights are given by $\langle jb|jb\rangle$ (*Cauchy–Schwarz*) or $\langle jb|bj\rangle$ (*Power–Pitzer*). If the total weight when finding b is zero (i.e. there are no spinorbitals with the required spin and symmetry or only the spinorbitals found as a has that spin and symmetry) or if the found b is already occupied, the spawn attempt is unsuccessful. Again, orbitals a and b were selected using their weights with the alias method.

The performance of the four excitation generators described in this subsection, *uniform Cauchy–Schwarz*, *heat bath Cauchy–Schwarz*, *uniform Power–Pitzer*, and *heat bath Power–Pitzer*, were then tested, using a chain of three water molecules in the cc-pVDZ basis,[68] whose molecular orbitals have been localized. The excitation generators all come with a low memory cost, which is $\mathcal{O}(M)$ temporarily or $\mathcal{O}(M^2)$ and all scale as $\mathcal{O}(M)$ in computational time. The distribution of $\frac{|\langle D_\mathbf{n}|\hat{H}|D_\mathbf{m}\rangle|}{p_{\mathrm{gen}}}$, which should ideally be constant, was compared for the four excitation generators. Figure 1 shows the histograms (excluding $\frac{|\langle D_\mathbf{n}|\hat{H}|D_\mathbf{m}\rangle|}{p_{\mathrm{gen}}} = 0$) with linear and logarithmic frequency scales. The *bottom* graph shows the all excitation generators have similar looking tails to both sides, the *heat bath Power–Pitzer* having the longest tail at big $\frac{|\langle D_\mathbf{n}|\hat{H}|D_\mathbf{m}\rangle|}{p_{\mathrm{gen}}}$. However, the number of events in bins above the maximum $\frac{|\langle D_\mathbf{n}|\hat{H}|D_\mathbf{m}\rangle|}{p_{\mathrm{gen}}}$ filled bin for the *uniform Power–Pitzer* excitation generator — which has the lowest maximum — is fewer than 100 events which is not significant relatively speaking so if not using initiator approximations there should not be a noticeable effect. The *top* graph demonstrates that the *heat bath Power–Pitzer* gives the sharpest peak and makes $\frac{|\langle D_\mathbf{n}|\hat{H}|D_\mathbf{m}\rangle|}{p_{\mathrm{gen}}}$ closest to a constant of
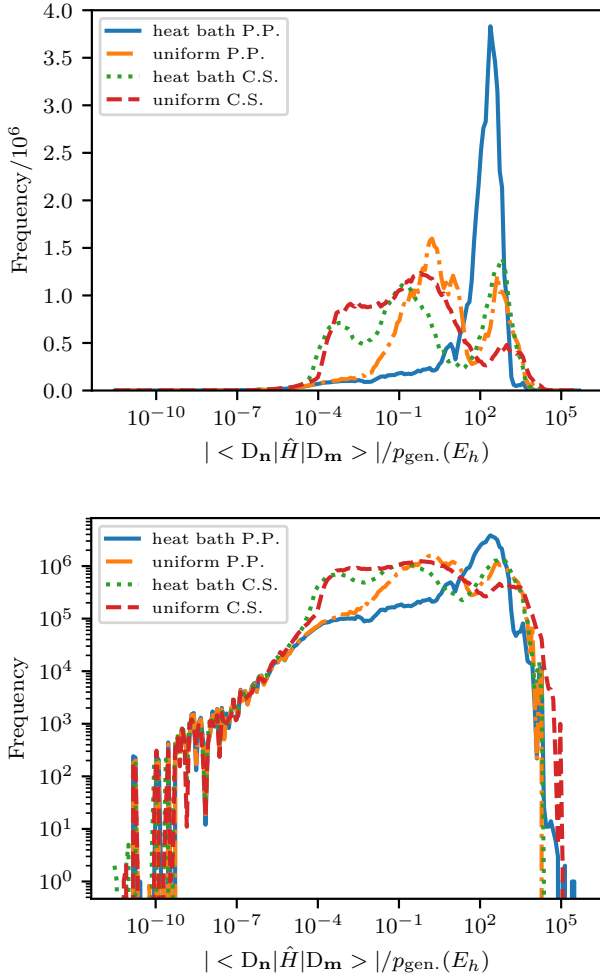
Figure 1: Comparison of the histograms of $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$ for the *Cauchy–Schwarz* (C.S.) and *Power–Pitzer* (P.P.) on-the-fly excitation generators. ij are either selected uniformly or using heat bath. The bin middles on the $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$ axis are used for the data points. The computational scaling of all excitation generators here is $\mathcal{O}(M)$. CCSD was performed on three water molecules in the cc-pVDZ basis using localized MOs. The values were logged for one Monte Carlo iteration. The size of the bins is logarithmic. *Bottom* graph took the log of the frequency whereas the *top* graph did not. They both show the same data. All of them were restarted from the same calculation and then equilibrated before taking data. $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}} = 0$ data is not shown which includes forbidden excitations. p_{single} was set to be the same when running which was corrected in post-processing to make the mean of finite $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$ for single and double excitations approximately coincide which did not change p_{single} values by more than about 30%.

the excitation generators. Only non-zero al-

Table 2: Fraction of allowed and fraction of non-zero allowed spawn events, both with respect to total number of spawn events. The latter represents the spawn events depicted in figure 1. *heat bath Cauchy–Schwarz* and *uniform Cauchy–Schwarz* and the *Power–Pitzer* excitation generators have been combined to *C.–S.* and *P.–P.* respectively. Individual data points have been rounded to the second decimal place. If a range is given they rounded to either value in the range.

	$\frac{\#\text{allowed}}{\#\text{total}}$ events	$\frac{\#\text{allowed\_non-zero}}{\#\text{total}}$ events
$C.–S.$	0.80	0.69–0.72
$P.–P.$	0.68–0.69	0.68–0.69

lowed events are shown in figure 1. Table 2 shows what fraction that is of the total number of events (second column) and what fraction of events are allowed which includes the allowed but zero $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$ events (first column). Both the *Cauchy–Schwarz* and the *Power–Pitzer* excitation generators have a similar fraction of non-zero allowed events. The *Power–Pitzer* excitation generators have more forbidden events but of those that are allowed, more are non-zero. A big source for forbidden events is the selection of b which is selected from the set of orbitals of required spin and symmetry which can be occupied. An event is then forbidden if b selected is occupied. Our implementation could be further improved by excluding occupied orbitals from that selection. In the results section we will let *heat bath Power–Pitzer* represent all these four excitation generators introduced in this subsection.

3.3 Pre-computed Power–Pitzer Excitation Generator

Even with their reduced memory requirements, the above excitation generators still add a considerable cost to calculations, and we seek a way to reduce this further. We now introduce an $\mathcal{O}(N)$ *Power–Pitzer* excitation generator, *heat bath Power–Pitzer ref.* , where N is the number of electrons. This can even be modified to be

$\mathcal{O}(N_{\text{ex.}})$ where $N_{\text{ex.}}$ is the number of electrons excited with respect to the reference if excitations instead of determinants were stored in our implementation. Within a routine coupled cluster calculation, the maximum $N_{\text{ex.}}$ does not depend on system size. This excitation generator combines advantages of *heat bath Power–Pitzer* where a bias check is not required beforehand (but is with the "original" *heat bath* excitation generator) and which has a significantly lower memory cost with those of the the lower computational scaling of the *heat bath* excitation generators, further improving upon them. We make use of the single-reference nature of coupled cluster where the reference determinant $|D_0\rangle$ is more important than any other determinant by pre-computing some weights based on the reference determinant. Pre-computed weights include heat bath and *Power–Pitzer* weights, for selecting the orbitals to excite from and to in a double excitation. Spinorbitals are first found by pretending the reference determinant is the determinant we are exciting from and are then mapped between the current determinant and the reference determinant when it is appropriate. The memory cost is $\mathcal{O}(M^2)$ while the computational cost when spawning is only the mapping of the reference $|D_0\rangle$ to the actual determinant $|D_m\rangle$ which is $\mathcal{O}(N)$. Since weights are based on one determinant, it is not costly to pre-calculate weights for single excitations as well. This is a considerable advantage over the on-the-fly *Power–Pitzer* and *heat bath* excitation generators that either do single excitations uniformly, exactly (which is costly) or partly based on double excitation weights.

In this algorithm, two frames of reference are considered. In the first frame, the reference frame, which is denoted by a prime, excitations are from the reference determinant, i.e. $|D_{m'}\rangle = |D_0\rangle$. In this frame, a double excitation would be $i'j' \to a'b'$. In the second frame, the simulation frame, the actual frame the calculation is in, excitations are from $|D_m\rangle$ and that excitation is $ij \to ab$. For selecting some orbitals, the weights of the orbitals in the reference frame are used and its spinorbitals are mapped to the simulation frame to find the actual excitation as explained in appendix B.

The following quantities for single excitations are pre-computed:

$$w_{i',s} = \sum_a \left(\frac{1}{n_{jb}} \sum_{j=j_{\text{occ.ref.}}, b=b_{\text{virt.ref.}}} \langle D_j^b| \hat{H} |D_{i'j}^{ab}\rangle \right),$$
(19)

where i' is an occupied orbital in the reference and the sum over a is over all orbitals with allowed excitation $i' \to a$. n_{jb} is $N(M-N)$. $|D_j^b\rangle$ differs from the reference determinant by the single excitation $j \to b$. We decided to not sum over single excitations from the reference as in the case of self-consistent field reference determinants, Brillouin's theorem would mean that the weights would be (close to) zero. We assume Brillouin's theorem when evaluating the weights. Assuming the system is single referenced, we might assume that a doubly excited determinant might be second most important after the reference determinant. The sum is therefore over all possible double excited determinants trying to connect to a determinant slightly closer to the reference via a single excitation. For selecting a,

$$w_{a=a_{\sigma,\text{sym.}},i,s} = \frac{1}{n_{jb}} \sum_{j=j_{\text{occ.ref.}}, b=b_{\text{virt.ref.}}} \langle D_j^b| \hat{H} |D_{ij}^{ab}\rangle,$$
(20)

is pre-computed where i is now an occupied orbital in the current determinant which will have been selected before $w_{a=a_{\sigma,\text{sym.}}i,s}$ is needed. Given that the current determinant is not known at this stage, this is pre-computed for any orbital i. a is then selected from the orbitals of allowed spin and symmetry for which $i \to a$ is valid. Alias tables are then pre-computed for $w_{i',s}$ and $w_{a=a_{\sigma,\text{sym.}}i,s}$.

When running the excitation generator, it is first decided whether a single or double excitation is attempted with probability p_{single} or $1 - p_{\text{single}}$ respectively. If a single excitation was chosen, i' is first selected in the reference frame from the occupied orbitals in the reference using the alias table constructed with weights $w_{i',s}$. i' is then mapped to the corresponding occupied orbital in the current determinant i in the simulation frame. Appendix B explains the mapping between these two frames in detail.

Once i is known, a is selected using the pre-computed alias table with $w_{a=a_{\sigma,\text{sym}.i,\text{s}}}$. Of course, a could be occupied. If that is the case, the excitation attempt was unsuccessful. Otherwise, $i \to a$ is found and the generation probability is

$$p_{\text{single}} \times \frac{w_{i',\text{s}}}{\sum_{i'=i'_{\text{occ.ref.}}} w_{i',\text{s}}} \times \frac{w_{a=a_{\sigma,\text{sym}.,i,\text{s}}}}{\sum_{a=a_{\sigma,\text{sym}.}} w_{a=a_{\sigma,\text{sym}.,i,\text{s}}}}. \tag{21}$$

For double excitations, four weight tables are pre-computed. For the selection of i and j, heat bath weights are pre-computed, assuming the reference determinant is fully occupied. Two orbitals i' and j' occupied in the reference are found and then mapped to the actual determinant that is occupied. For the virtual orbitals a and b, alias tables based on Power–Pitzer weights are pre-calculated for *all* spinorbitals. Before selecting a, the actual i is known and can be substituted into pre-computed weights $\sqrt{|\langle ia|ai\rangle|}$ to find a. The memory cost is $\mathcal{O}(M^2)$. No mapping is necessary for a and b. Again, if a or b are occupied or b is equal to a or if there is not suitable orbital for b, the spawn attempt was unsuccessful. Double excitations with this excitation generator are explained in more detail in appendix C.

Overall, this is an excitation generator that is both weighted and can scale as $\mathcal{O}(N_{\text{ex.}})$ in CCMC which does not scale with system size. In FCIQMC the scaling is still low, $\mathcal{O}(N)$. The memory cost is also relatively small, $\mathcal{O}(M^2)$.

4   Results and Discussion

To compare the effectiveness of the excitation generators discussed, water chains were then studied in a cc-pVDZ basis set[68] whose MOs have been localized. Figure 2 shows a histogram of $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$ for three waters with the four uniform excitation generators, the *heat bath Power–Pitzer* excitation generator (which had the sharpest peak out of the $\mathcal{O}(M)$/on-the-fly excitation generators), the *heat bath Power–Pitzer ref.* and the two *heat bath* excitation gen-

erators that do not suffer from bias. The "original" *heat bath* excitation generator was rejected by our bias test as it was not clear whether all allowed single excitations can be created. Considering a logarithmic scale in $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$, the *top* graph in figure 2 clearly shows that the uniform excitation generators produce a bigger spread in $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$ than weighted excitation generators (*Power–Pitzer* or *heat bath*).

The *heat bath* excitation generators produce the sharpest peak. The *heat bath uniform singles* excitation generator, that samples single excitations uniformly, shares the main peak with the *heat bath exact singles* excitation generator, that samples single excitations exactly, but has a larger spread around that peak caused by the uniform sampling of single excitations. The *heat bath exact singles* excitation generator produces two sharp peaks, both containing data from single excitations which were treated exactly here. The reason why this is not one sharp peak is that in an ideal case

$$p_{\text{gen.}} = \left| \frac{\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle}{\sum_{\mathbf{n}} \langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle} \right| \tag{22}$$

which means that

$$\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}} \approx \frac{1}{|\sum_{\mathbf{n}} \langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|} \tag{23}$$

in the case of an ideal excitation generator. This quantity depends on $|D_{\mathbf{m}}\rangle$ and can therefore not be a constant in general unless the selection step in the CCMC or FCIQMC algorithm is adapted as well. Both *heat bath* excitation generators here have a large memory scaling ($\mathcal{O}(M^4)$) and *heat bath exact singles* which produces the sharpest peak in the histogram has a computational scaling of $\mathcal{O}(MN^2)$ (FCIQMC and in this implementation) or $\mathcal{O}(MNN_{\text{ex.}})$ (ideal implementation CCMC) which makes the *heat bath exact singles* excitation generator not practical.

The main peak that the two *Power–Pitzer* excitation generators produce is wider than with the *heat bath* excitation generators but it is significantly more compact that what the uniform excitation generators give. The *heat*
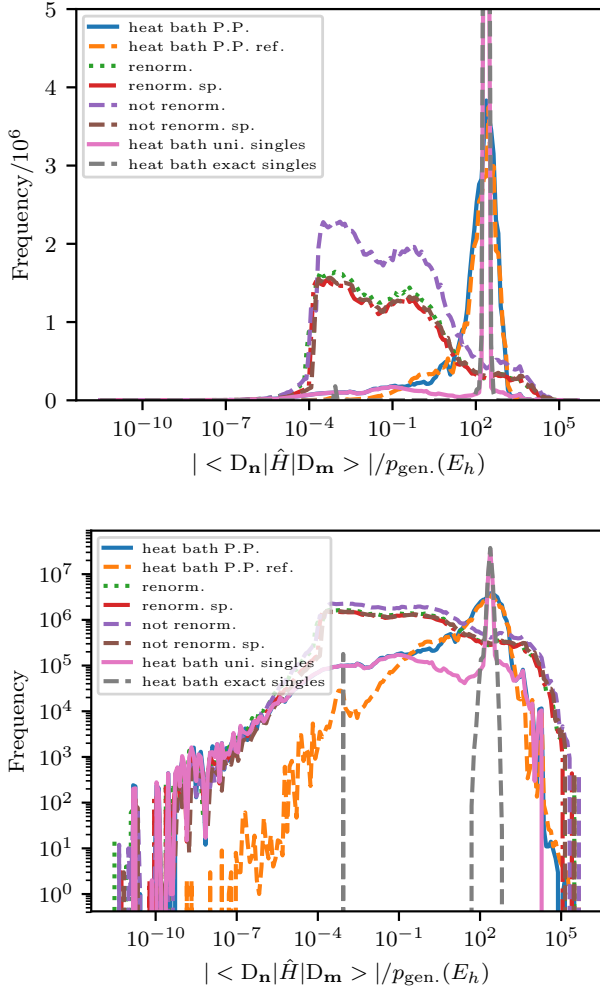
Figure 2: Comparison of the histograms of $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$ for various excitation generators. The bin middles on the $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$ axis are used for the data points. CCSD was performed on three water molecules in the cc-pVDZ basis using localized MOs. The values were logged for one Monte Carlo iteration. The size of the bins is logarithmic. *Bottom* graph took the log of the frequency whereas the *top* graph did not. They both show the same data. The frequency axis in the case is truncated in the *top* graph. Most of them were restarted from the same calculation and then equilibrated before taking data. *heat bath exact singles* was restarted from an equilibrated *heat bath uniform singles* but not equilibrated since it is very slow. $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}} = 0$ data is not shown which includes forbidden excitations. p_{single} was set to be the same when running which was corrected in post-processing to make the mean of finite $\frac{|\langle D_{\mathbf{n}}|\hat{H}|D_{\mathbf{m}}\rangle|}{p_{\text{gen}}}$ for single and double excitations approximately coincide which did not change p_{single} values by more than 30%.

bath Power–Pitzer ref. excitation generator has a shorter tail on the low end but a slightly wider tail on the higher end. It has fewer than 250 events in bins with bigger $\frac{|\langle D_{\mathbf{m}}|\hat{H}|D_{\mathbf{n}}\rangle|}{p_{\text{gen}}}$ than the highest bin that has an event with the *heat bath uniform singles* excitation generator. The *heat bath Power–Pitzer* excitation generator has fewer than 90 events above the bin with highest $\frac{|\langle D_{\mathbf{m}}|\hat{H}|D_{\mathbf{n}}\rangle|}{p_{\text{gen}}}$ in the *heat bath uniform singles* case. The number of finite

Table 3: Fraction of non-zero allowed spawn events, both with respect to total number of spawn events. The latter represents the spawn events depicted in figure 2. The *renorm.* and *renorm. spin* have been combined to *renorm.* and similarly for *not. renorm.*. *P.–P.* means *Power–Pitzer* and *heat b.* is *heat bath*. Individual data points have been rounded to the second decimal place. If a range is given they rounded to either value in the range.

	$\frac{\text{\#allowed non−zero}}{\text{\#total}}$ events
heat b. P.–P. ref.	0.66–0.67
heat b. P.–P.	0.69
heat b. uniform singles	0.72
heat b. exact singles	0.72
renorm.	0.69–0.72
not. renorm.	0.54–0.57

$\frac{|\langle D_{\mathbf{m}}|\hat{H}|D_{\mathbf{n}}\rangle|}{p_{\text{gen}}}$, allowed events are shown in table 3. The weighted excitation generators have similar fractions of allowed non-zero events and the *heat bath Power–Pitzer ref.* excitation generator has the lower computational scaling compared to *heat bath Power–Pitzer* and the *heat bath uniform singles* excitation generator, at least in the case of CCMC. It also does not have the prohibitively large memory scaling of the *heat bath uniform singles* excitation generator.

Next, we move away from abstract performance considerations and consider how the different excitation generators affect the the efficiency (as described by Holmes et al.[20]), inefficiency,[69] and the position of the shoulder[7] which are all measures of the difficulty of the calculation. The efficiency η is defined as $\eta = 1/(\sigma_E^2 T)$, where σ_E is the statistical error in the energy (here projected energy) and T is the

computational time taken to achieve error bar σ_E. In our case here, T was estimated by the CPU time, that sums over OpenMP threads, as determined by the parent MPI process. It is then multiplied by the number of MPI processes. It is therefore to be treated as an approximation. T is the sum of individual times for blocks of iterations and only times of iterations actually used by the reblocking analysis are summed up. We have found T to be highly dependent on implementation so η must be considered carefully. We also consider the (theoretical) algorithmic computational scaling in mind and the inefficiency a as defined by Vigor et al.,[69] $a = \sigma_E \sqrt{\delta \tau N_{\text{it.}} \langle N_{\text{tot.}} \rangle}$ where $N_{\text{it.}}$ is the number of iterations considered in the blocking analysis and $\langle N_{\text{tot.}} \rangle$ is the mean number of Monte Carlo particles. When estimating the error in the efficiency and inefficiency, we ignore the correlation in the numerator and denominator of the $E_{\text{proj.}}$.

4.1 Coupled Cluster Monte Carlo

All coupled cluster calculations are non-initiator.[7,13] Figure 3 shows the efficiency and inefficiency for chains of two or three waters in the cc-pVDZ basis performing CCSD/CCSDT with localized or canonical molecular orbitals. CCSDT was only run on the weighted excitation generators. When localization has not been applied, i.e. our canonical CCSD run, symmetry has been ignored as it also does not exist in the system with localized orbitals. The systems to study where chosen not to be too large to get small enough error bars on efficiency and inefficiency. However, the basis set could not be too small since the *heat bath uniform singles* and the *heat bath Power–Pitzer ref.* excitation generators assume that the number of occupied orbitals is small relative to the number of total orbitals, which reflects a realistic calculation. Note that while all of the four types of calculations were run with the same number of MPI processes and OpenMP threads for the different excitation generators, these numbers varied between the types of calculations [9]. The

heat bath exact singles excitation generator is so slow that it was not possible to take sufficient data with it to produce results.

We now discuss the trends shown in figure 3:

- *System size:* The overall trend is that the weighted excitation generators are more efficient and less inefficient than the uniform ones. This becomes more noticeable in the larger system. As expected, modelling three waters is less efficient and more inefficient than two, the difference being more distinct with the uniform excitation generators.

- *Coupled cluster level:* When raising the excitation level to CCSDT, which we did for the weighted excitation generators, the inefficiency increases and the efficiency decreases slightly compared to the CCSD calculation. This is to be expected as the Hilbert space to cover increases. All three weighted excitation generators are affected.

- *Localization:* Using orbitals that have not been localized does not seem to affect the efficiency of the uniform excitation generators and *heat bath uni. singles*. The *heat bath Power–Pitzer (ref.)* excitation generators show a decline in efficiency and increase in inefficiency. In fact, they seem to drop to a similar efficiency level as the uniform excitation generators, *heat bath Power–Pitzer ref.* still being slightly more efficient. The inefficiency means of the *heat bath Power–Pitzer (ref.)* excitation generators are lower than the ones from the uniform excitation generators, even though the error bars overlap. We expect localization to primarily to affect the weighted excitation generators as, in a double excitation, the weights in the *heat bath uni. singles* are calculated as a sum of Coulomb and — if the spins

---

[9]To be precise: The dimer calculations were done with 2, 1 and 8 MPI processes for the CCSD localized,

CCSD canonical and CCSDT localized calculation respectively. The trimer calculation has been done with 4 MPI processes. The CCSD canonical calculation used 24 OpenMP threads for its MPI process, all the others 12 OpenMP threads per MPI process.
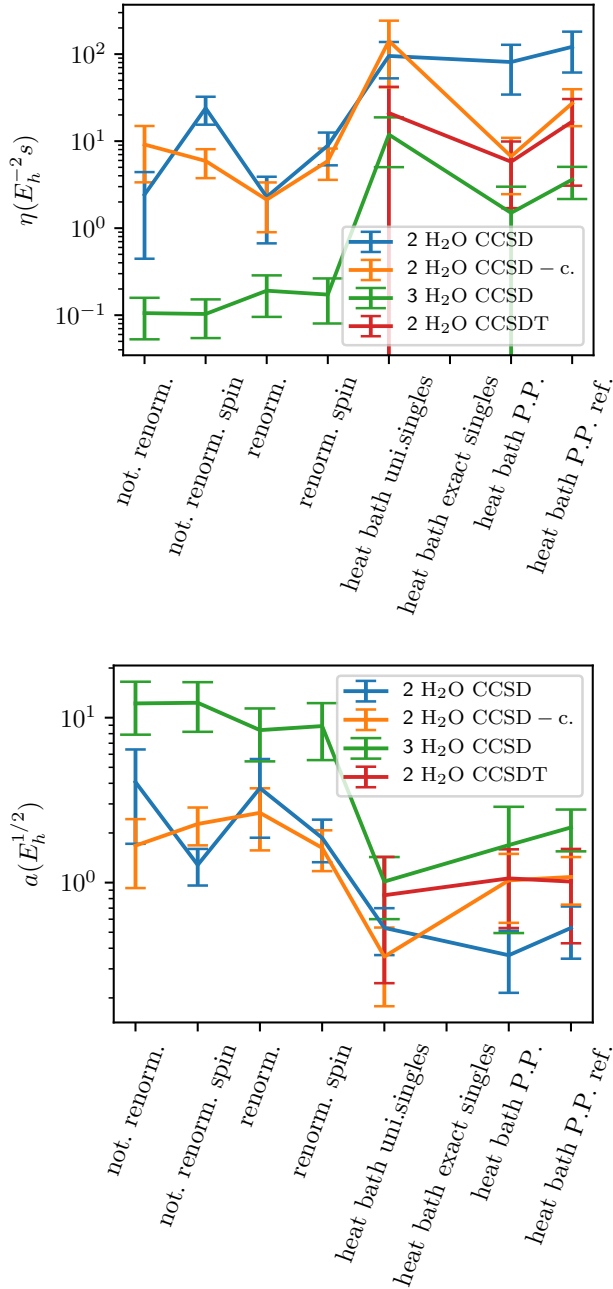
Figure 3: Efficiency η (*top*) and inefficiency a (*bottom*) for chains of two or three water molecules in a cc-pVDZ basis run with CCSD/CCSDT using localized/canonical ('-c.') MOs. Error bars neglect the covariance between numerator and denominator errors in the projected energy. The *heat bath exact singles* excitation generator was too slow for data to be taken. The different excitation generators were run under the same conditions with the same time step etc. Only the target population was varied between the calculations. The starting iteration for *heat bath P.P.* was found such that three reblocks could be used in the trimer calculation. The number of reblocks was raised for *heat bath P.P. ref* for the CCSDT calculation by taking the result of the previous reblocking iteration for the projected energy. The shift and the projected energy disagreed by more than 2 standard errors in the canonical CCSD run with *heat bath P.P. ref.* and *(not.) renorm. spin* and in the CCSDT run with *heat bath P.P. ref.* excitation generators.

are parallel — exchange integrals whereas the *heat bath Power–Pitzer (ref.)* excitation generators only use exchange integral weights. Coulomb integrals decay as the inverse of the distance but exchange integrals are more affected by the localization. This explains why the *heat bath Power–Pitzer (ref.)* excitation generator efficiencies are more strongly affected by localization.

The *heat bath uniform singles* excitation generator performs best out of the weighted ones which is expected due to the same low computational scaling as *heat bath Power–Pitzer ref.* which is more favourable than *heat bath Power–Pitzer* while using well approximated weights for double excitations. *heat bath Power–Pitzer ref.* also seems to have higher efficiencies than *heat bath Power–Pitzer*, likely due to the better computational scaling and possibly the more accurate treatment of single excitations, and this might yet be improved by a better *heat bath Power–Pitzer ref.* implementation which scales as $\mathcal{O}(N_{ex.})$ rather than $\mathcal{O}(N)$ computationally.

Next, we consider shoulder heights with CCSDT on two water molecules with localized orbitals. Shoulder heights indicate approximately the minimum number of particles needed in the simulation. Figure 4 shows shoulder plots where the difference in shoulder positions between the excitation generators is very clear. The weighted excitation generators again perform best. Their shoulders are significantly lower than those of uniform excitation generators, by a factor of just under 2. Of those studied, the *heat bath Power–Pitzer ref.* excitation generator has the lowest shoulder.

With localized orbitals, the weighted excitation generators all perform better than the uniform ones. The *heat bath Power–Pitzer ref.* excitation generator can scale independently of system size computationally which puts it at a clear advantage over the *heat bath Power–Pitzer* excitation generator. It also has a reduced memory scaling when comparing it to the *heat bath* excitation generators which is significant at bigger systems.

Figure 4: Shoulder plots for two localized waters in a cc-pVDZ basis with CCSDT with various excitation generators. P.P. stands for *Power–Pitzer*, h.b. for *heat bath* and sp. for *spin*. The different excitation generators were run under similar conditions with the same time step etc. The weighted excitations generators started varying the shift after a total population of 20 million whereas the uniform ones did not vary the shift. The vertical lines represent the "shoulder height", the position of the maximum plus/minus of a standard deviation. To determine the shoulder position, the mean and standard error of the mean of the 10 highest data points were taken.[7]

4.2 Full Configuration Interaction Quantum Monte Carlo

Next, we turn to FCIQMC. The water chain with two waters in cc-pVDZ basis with localized MOs was considered with initiator FCIQMC. The (in–)efficiencies were determined at one point in the initiator curve (total population against energy). All calculations were started with the same parameters, which included the population at which the shift started varying, and so the eventual equilibrated population of the system is dependent upon the excitation generator. Blooms did happen. For uniform excitation generators it was over 10^7 particles, for the weighted ones 5.6×10^6. Use of a larger population may lead to a decrease in measured inefficiency,[69] so the results from the uniform excitation generators should be regarded as lower bounds for inefficiency.

Figure 5 shows the efficiency and inefficiency for that system with the particle populations $N_{tot.}$ explicitly indicated. The weighted excitation generators perform comparably among themselves and all outperform the uniform ones. *heat bath Power–Pitzer ref.* and *heat bath uniform singles* excitation generators both scale linearly in the number of electrons when using FCIQMC. This study has been done on a single point in the initiator curve and we did not investigate whether the behaviour of the initiator curve changed which can affect number of particles needed for convergence. Holmes et al.[20] describe ways to reduce the memory cost by considering spins (we just store zeroes instead of considering the spin when selecting) or by not storing all the weight to select b for example. We have used double precision for the weights. However, even if our implementation is not optimal, it is clear that the *heat bath* excitation generators hit a memory ceiling with big systems significantly earlier than the *heat bath Power–Pitzer ref.* excitation generators. Also, as mentioned earlier, our *heat bath Power–Pitzer* excitation generator implementation could be improved by making sure b is only selected from virtual orbitals. However, even with a more ideal code, the computational scaling of $\mathcal{O}(M)$ remains which becomes pro-
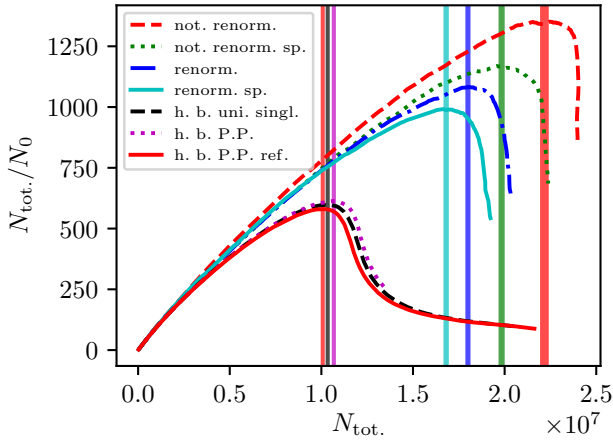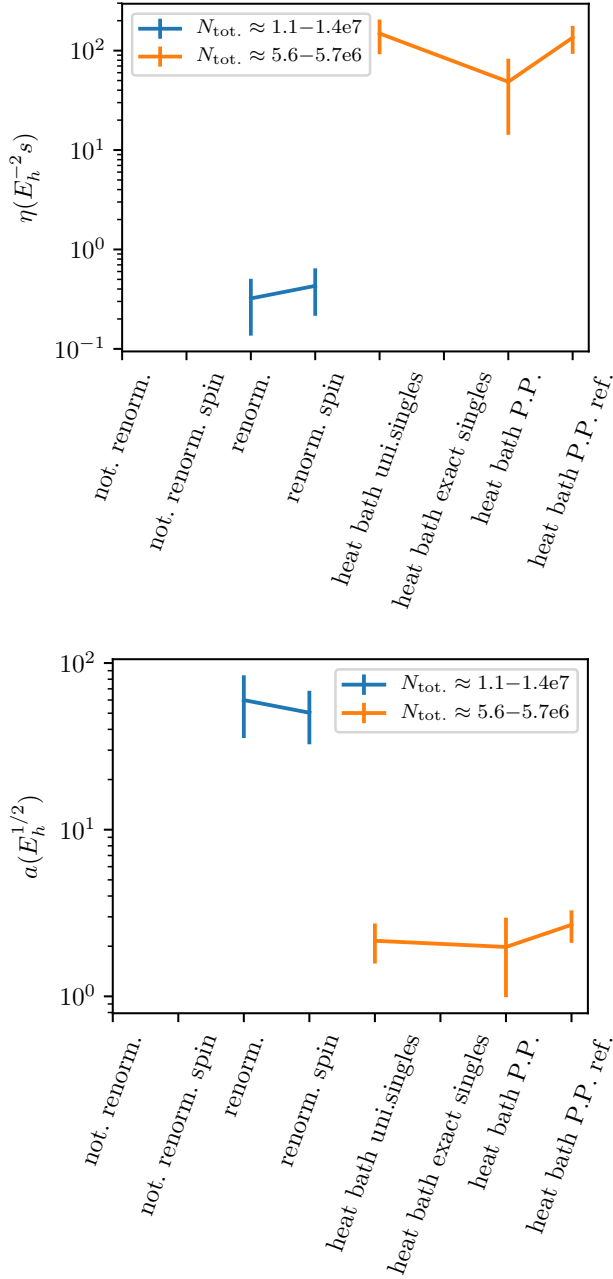
15

Figure 5: Efficiency η (*top*) and inefficiency a (*bottom*) for a chain of two water molecules in cc-pVDZ basis using localized MOs run with initiator FCIQMC with approximate MC particle populations indicated. Error bars neglect the covariance between numerator and denominator errors in the projected energy and are over-estimates. The *heat bath exact singles* excitation generator was too slow for data to be taken. The different excitation generators were run under the same conditions with the same time step etc. The spawning arrays of the *not. renorm.* excitation generators ran out of memory so the space to store the spawned walkers would need to be increased for those results.

hibitive in large systems.

This suggests that *heat bath Power–Pitzer ref.* is an efficient excitation generator with a low shoulder that can be used in CCMC and FCIQMC as a weighted excitation generator with low computational and memory cost.

4.3   Practical Advice

As long as the memory allows it, it makes sense to make use of the *heat bath* (if no bias is present) or *heat bath uniform singles* excitation generator for FCIQMC as they have the same computational scaling as the *heat bath Power–Pitzer ref.* excitation generator and a better scaling than the *heat bath Power–Pitzer* excitation generator. The results in this paper suggest that they are also at least as efficient as the *heat bath Power–Pitzer ref.* excitation generator. For CCMC with an implementation where the *heat bath Power–Pitzer ref.* excitation generator has a scaling of $\mathcal{O}(N_{\text{ex.}})$, the *heat bath Power–Pitzer ref.* excitation generator using localized orbitals is suitable.

As the system size becomes more substantial, the *heat bath* excitation generators will fail due to memory requirements. In that case, the *heat bath Power–Pitzer ref.* excitation generator should be considered, ideally with localized orbitals.

5   Conclusion

We have shown that especially when using localized orbitals the *heat bath Power–Pitzer ref.* excitation generator combines the advantages of *heat bath* excitation generators, which are relatively fast and use good weights but struggle with a significant memory cost and a possible bias, and the excitation generators that approximate heat bath weights by inequalities which are calculated on-the-fly reducing the memory scaling but scaling prohibitively computationally in big systems. The *heat bath Power–Pitzer ref.* excitation generator has at worst a low computational order and can be implemented with computational cost independent of system size in coupled cluster with a low memory cost.

A Further Uniform Excitation Generators

In the case of a double excitation, Hamiltonian matrix elements tend to be bigger if i and j do not have parallel spins. This is because following Slater-Condon rules,[65,66] the Hamiltonian matrix element is reduced to a sum of two terms of opposite sign in the case of parallel spins ($\langle ij|ab\rangle - \langle ij|ba\rangle$) and one term if the spins are not parallel ($\langle ij|ab\rangle$). It might therefore be advisable to select anti-parallel spin electrons with a greater probability than parallel ij. Alavi, Booth and others[19][10] had the idea of determining whether spins are antiparallel or parallel first when selecting i and j. The *no. renorm. spin* and *renorm. spin* excitation generators are modifications of *no. renorm.* and *renorm.* excitation generators, where instead of finding i and j as a pair from the set of occupied orbitals, it is first decided whether they should have parallel spins or not. With probability p_{parallel}, ij are either selected as a pair from the set of occupied α (probability $\frac{N_\alpha}{N}$) or from the set of occupied

---

¹⁰Personal Communication with Ali Alavi and Pablo López Ríos. This is also implemented in NECI https://github.com/ghb24/NECI_STABLE.

β orbitals (probability $1 - \frac{N_\alpha}{N} = \frac{N_\beta}{N}$) where N_α and N_β are the number of α and β electrons respectively. This can lead to forbidden excitations followed by failed spawning attempts if there is only one electron of one type of spin. Here, p_{parallel} is set as the fraction of H_{ijab} where i and j have parallel spins.

B Mapping spinorbitals in *heat bath Power–Pitzer ref.* excitation generator

In HANDE, there is a list of orbitals that are occupied in the reference, usually approximately ordered by one electron energies, and there is an equivalent ordered list with orbitals occupied in the current determinant $|D_{\mathbf{m}}\rangle$. The localized orbitals here were ordered by approximate orbital energies, given by the expectation value of the Fock operator. Every time $|D_{\mathbf{m}}\rangle$ is changed, two new (energy ordered) lists RD and CD are created, one (RD) containing all orbitals that are occupied in the reference but not in $|D_{\mathbf{m}}\rangle$ and another list (CD) of the same size with all orbitals occupied in $|D_{\mathbf{m}}\rangle$ but not the reference determinant. Orbitals with the same positions in these two lists are made to have the same spin by swapping orbitals in the list CD if necessary. If necessary, orbitals are translated by a one-to-one mapping between these two lists. If i' is not only occupied in the reference but in $|D_{\mathbf{m}}\rangle$ as well, $i' = i$. If not, the position i' has in list RD is translated to the orbital with the same position in list CD. Figure 6 shows the translation of i and j in a double excitation in the two frames of reference pictorially. Note that this is the only part of the excitation generator that is not $\mathcal{O}(1)$ but $\mathcal{O}(N)$, mainly arising due to the creation of the two lists. The computational cost is reduced to $\mathcal{O}(1)$ if a determinant is reused. Alternatively, if, as mentioned previously, each excitor is not represented by a determinant but rather the lists RD or CD from the beginning the scaling is reduced to $\mathcal{O}(N_{\text{ex.}})$ which is the cost of finding the correct mapping from one list to the other.
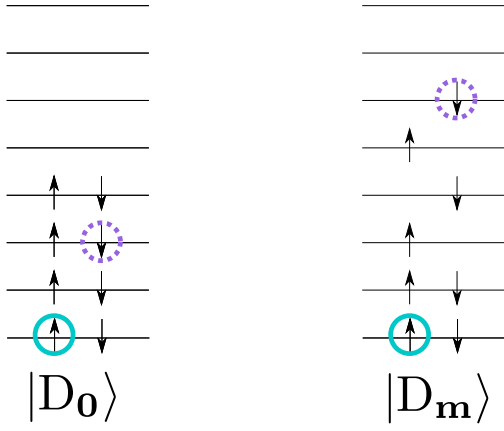
Figure 6: Selecting i and j with *heat bath Power–Pitzer ref.* excitation generator for a double excitation. First i' is selected, occupied in the reference determinant $|D_0\rangle$ and translated to i, occupied in the current determinant, $|D_m\rangle$. i' and i are shown with light blue solid circles. In this case, $i' = i$. Then j' is found and translated to j. As j' is not occupied in $|D_m\rangle$, it is mapped to the next orbital of the same spin occupied in $|D_m\rangle$ but not in $|D_0\rangle$. j' and j are shown with dashed purple circles. Here $j' \neq j$.

C Details of double excitations in the *heat bath Power–Pitzer ref.* excitation generator

Again, orbitals $i'j'$ are part of the reference frame, where the reference determinant is occupied, and ij are the equivalent spinorbitals in the actual frame, where the actual determinant we are exciting from is occupied. $i'j'$ are first found in the reference frame using heat bath weights and then they are mapped to the actual frame as described in appendix B. ab are found with Power–Pitzer weights in the actual frame. All weights are pre-computed. This appendix describes the details of generating the double excitation. For i', the pre-computed weights are

$$w_{i',\mathrm{d}} = \sum_{j'=j_{\mathrm{occ.ref.},\neq i'},a_{\neq\{i',j'\}},b_{\neq\{i',j',a\}}} H_{i'j'ab} \quad (24)$$

i' is selected from the set of occupied orbitals in the reference with a sum over j', the set of occupied orbitals in the reference other than i'. a and b out of the set of all orbitals (not just virtual) are summed over, provided they don't equal i', j' or each other. For j',

$$w_{j'i,\mathrm{d}} = \sum_{a_{\neq\{i,j'\}},b_{\neq\{i,j',a\}}} H_{ij'ab} \quad (25)$$

is pre-calculated which is of memory scaling order $\mathcal{O}(NM)$. For both $w_{i',\mathrm{d}}$ and $w_{j'i,\mathrm{d}}$, a minimum weight is set in case the total weight for selecting i' or j' respectively in the reference frame is zero but selecting the equivalent i and j in the simulation frame would be allowed.

To select a and b, Power–Pitzer weights are pre-calculated. For a,

$$w_{a,i,\mathrm{d}} = \sqrt{|\langle ia|ai\rangle|} \quad (26)$$

where $w_{a,i,\mathrm{d}}$ is zero if $i = a$. ia are from the set of all spinorbitals and a is restricted to the set of the same spin as i. The memory cost is simply $\mathcal{O}(M^2)$. Similarly, for b

$$w_{b,j,\mathrm{sym.,d}} = \sqrt{|\langle jb|bj\rangle|} \quad (27)$$

where $w_{b,j,\mathrm{d}} = 0$ if $b = j$ and b is from the set of all spinorbitals with the same spin as j. $w_{b,j,\mathrm{d}}$ are arranged in such a way that b's of the required symmetry later can readily be looked up. Alias tables for all these weights for single and double excitations are pre-computed.

In the case of a double excitation, first i', an occupied orbital in the reference frame, is selected using $w_{i',\mathrm{d}}$. $i' \rightarrow i$ is mapped to an occupied orbital i in the simulation frame if required. Then, j' is found using the pre-computed alias table for $w_{j'i,\mathrm{d}}$ and map $j' \rightarrow j$ if needed. i and j are ordered so that j has a higher or equal index in the determinant list as i. Using i and $w_{a,i,\mathrm{d}}$, a is found using pre-computed alias tables out of all spinorbitals with the same spin as i. If a is occupied, the spawn attempt was unsuccessful. The symmetry that b should have is then determined and using the pre-calculated alias tables for $w_{b,j,\mathrm{sym.,d}}$ which give us a b of the correct symmetry (and spin), b is found from the set of all

spinorbitals with required spin and symmetry. Again, if b is occupied or equal to a or if there is not suitable orbital for b, the spawn attempt was unsuccessful.

References

(1) Coester, F.; Kümmel, H. Short-range correlations in nuclear wave functions. *Nucl. Phys.* **1960**, *17*, 477–485.

(2) Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966**, *45*, 4256–4266.

(3) Čížek, J.; Paldus, J. Correlation problems in atomic and molecular systems III. Rederivation of the coupled-pair many-electron theory using the traditional quantum chemical methodst. *Int. J. Quantum Chem.* **1971**, *5*, 359–379.

(4) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291–352.

(5) Lee, T. J.; Scuseria, G. E. *Quantum Mech. Electron. Struct. Calc. with Chem. Accuracy*; Springer Netherlands: Dordrecht, 1995; pp 47–108.

(6) Thom, A. J. W. Stochastic Coupled Cluster Theory. *Phys. Rev. Lett.* **2010**, *105*, 263004.

(7) Spencer, J. S.; Thom, A. J. W. Developments in stochastic coupled cluster theory: The initiator approximation and application to the uniform electron gas. *J. Chem. Phys.* **2016**, *144*, 084108.

(8) Franklin, R. S. T.; Spencer, J. S.; Zoccante, A.; Thom, A. J. W. Linked coupled cluster Monte Carlo. *J. Chem. Phys.* **2016**, *144*, 044111.

(9) Scott, C. J. C.; Thom, A. J. W. Stochastic coupled cluster theory: Efficient sampling of the coupled cluster expansion. *J. Chem. Phys.* **2017**, *147*, 124105.

(10) Spencer, J. S.; Neufeld, V. A.; Vigor, W. A.; Franklin, R. S. T.; Thom, A. J. W. Large Scale Parallelization in Stochastic Coupled Cluster. *arXiv [physics.chem-ph]* **2018**,

(11) Neufeld, V. A.; Thom, A. J. W. A study of the dense uniform electron gas with high orders of coupled cluster. *J. Chem. Phys.* **2017**, *147*, 194105.

(12) Booth, G. H.; Thom, A. J. W.; Alavi, A. Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space. *J. Chem. Phys.* **2009**, *131*, 054106.

(13) Cleland, D.; Booth, G. H.; Alavi, A. Communications: Survival of the fittest: Accelerating convergence in full configuration-interaction quantum Monte Carlo. *J. Chem. Phys.* **2010**, *132*, 041103.

(14) Spencer, J. S.; Blunt, N. S.; Foulkes, W. M. The sign problem and population dynamics in the full configuration interaction quantum Monte Carlo method. *J. Chem. Phys.* **2012**, *136*, 054110.

(15) Foulkes, W. M. C.; Mitas, L.; Needs, R. J.; Rajagopal, G. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.* **2001**, *73*, 33–83.

(16) Booth, G. H.; Cleland, D.; Thom, A. J. W.; Alavi, A. Breaking the carbon dimer: The challenges of multiple bond dissociation with full configuration interaction quantum Monte Carlo methods. *J. Chem. Phys.* **2011**, *135*, 084104.

(17) Cleland, D.; Booth, G. H.; Overy, C.; Alavi, A. Taming the First-Row Diatomics: A Full Configuration Interaction Quantum Monte Carlo Study. *J. Chem. Theory Comput.* **2012**, *8*, 4138–4152.

(18) Daday, C.; Smart, S.; Booth, G. H.; Alavi, A.; Filippi, C. Full Configuration Interaction Excitations of Ethene and Butadiene: Resolution of an Ancient Question. *J. Chem. Theory Comput.* **2012**, *8*, 4441–4451.

(19) Booth, G. H.; Smart, S. D.; Alavi, A. Linear-scaling and parallelisable algorithms for stochastic quantum chemistry. *Mol. Phys.* **2014**, *112*, 1855–1869.

(20) Holmes, A. A.; Changlani, H. J.; Umrigar, C. J. Efficient Heat-Bath Sampling in Fock Space. *J. Chem. Theory Comput.* **2016**, *12*, 1561–1571.

(21) Sharma, S.; Yanai, T.; Booth, G. H.; Umrigar, C. J.; Chan, G. K.-L. Spectroscopic accuracy directly from quantum chemistry: Application to ground and excited states of beryllium dimer. *J. Chem. Phys.* **2014**, *140*, 104112.

(22) Veis, L.; Antalík, A.; Legeza, Ö.; Alavi, A.; Pittner, J. Full configuration interaction quantum Monte Carlo benchmark and multireference coupled cluster studies of tetramethyleneethane diradical. *arXiv [physics.chem-ph]* **2018**,

(23) Samanta, P. K.; Blunt, N. S.; Booth, G. H. Response Formalism within Full Configuration Interaction Quantum Monte Carlo: Static Properties and Electrical Response. *J. Chem. Theory Comput.* **2018**, *14*, 3532–3546.

(24) Shepherd, J. J.; Grüneis, A.; Booth, G. H.; Kresse, G.; Alavi, A. Convergence of many-body wave-function expansions using a plane-wave basis: From homogeneous electron gas to solid state systems. *Phys. Rev. B* **2012**, *86*, 035111.

(25) Shepherd, J. J.; Booth, G.; Grüneis, A.; Alavi, A. Full configuration interaction perspective on the homogeneous electron gas. *Phys. Rev. B* **2012**, *85*, 081103.

(26) Shepherd, J. J.; Booth, G. H.; Alavi, A. Investigation of the full configuration interaction quantum Monte Carlo method using homogeneous electron gas models. *J. Chem. Phys.* **2012**, *136*, 244101.

(27) Booth, G. H.; Grüneis, A.; Kresse, G.; Alavi, A. Towards an exact description of electronic wavefunctions in real solids. *Nature* **2013**, *493*, 365–370.

(28) Schwarz, L. R.; Booth, G. H.; Alavi, A. Insights into the structure of many-electron wave functions of Mott-insulating antiferromagnets: The three-band Hubbard model in full configuration interaction quantum Monte Carlo. *Phys. Rev. B* **2015**, *91*, 045139.

(29) Luo, H.; Alavi, A. Combining the Transcorrelated Method with Full Configuration Interaction Quantum Monte Carlo: Application to the Homogeneous Electron Gas. *J. Chem. Theory Comput.* **2018**, *14*, 1403–1411.

(30) Ruggeri, M.; Ríos, P. L.; Alavi, A. Correlation energies of the high-density spin-polarized electron gas to meV accuracy. *arXiv[cond-mat.str-el]* **2018**,

(31) Booth, G. H.; Chan, G. K.-L. Communication: Excited states, dynamic correlation functions and spectral properties from full configuration interaction quantum Monte Carlo. *J. Chem. Phys.* **2012**, *137*, 191102.

(32) Ten-no, S. Stochastic determination of effective Hamiltonian for the full configuration interaction solution of quasi-degenerate electronic states. *J. Chem. Phys.* **2013**, *138*, 164126.

(33) Humeniuk, A.; Mitrić, R. Excited states from quantum Monte Carlo in the basis of Slater determinants. *J. Chem. Phys.* **2014**, *141*, 194104.

(34) Blunt, N. S.; Smart, S. D.; Booth, G. H.; Alavi, A. An excited-state approach

within full configuration interaction quantum Monte Carlo. *J. Chem. Phys.* **2015**, *143*, 134117.

(35) Blunt, N. S.; Booth, G. H.; Alavi, A. Density matrices in full configuration interaction quantum Monte Carlo: Excited states, transition dipole moments, and parallel distribution. *J. Chem. Phys.* **2017**, *146*, 244105.

(36) Deustua, J. E.; Shen, J.; Piecuch, P. Converging High-Level Coupled-Cluster Energetics by Monte Carlo Sampling and Moment Expansions. *Phys. Rev. Lett.* **2017**, *119*, 223003.

(37) Petruzielo, F. R.; Holmes, A. A.; Changlani, H. J.; Nightingale, M. P.; Umrigar, C. J. Semistochastic Projector Monte Carlo Method. *Phys. Rev. Lett.* **2012**, *109*, 230201.

(38) Blunt, N. S.; Smart, S. D.; Kersten, J. A. F.; Spencer, J. S.; Booth, G. H.; Alavi, A. Semi-stochastic full configuration interaction quantum Monte Carlo: Developments and application. *J. Chem. Phys.* **2015**, *142*, 184107.

(39) Kersten, J. A. F.; Booth, G. H.; Alavi, A. Assessment of multireference approaches to explicitly correlated full configuration interaction quantum Monte Carlo. *J. Chem. Phys.* **2016**, *145*, 054117.

(40) Smart, S. D.; Booth, G. H.; Alavi, A. Excitation generation in full configuration interaction quantum Monte Carlo based on Cauchy-Schwarz distributions. *unpublished*

(41) Blunt, N. S. Communication: An efficient and accurate perturbative correction to initiator full configuration interaction quantum Monte Carlo. *J. Chem. Phys.* **2018**, *148*, 221101.

(42) Holmes, A. A.; Tubman, N. M.; Umrigar, C. J. Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 3674–3680.

(43) Sharma, S.; Holmes, A. A.; Jeanmairet, G.; Alavi, A.; Umrigar, C. J. Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory. *J. Chem. Theory Comput.* **2017**, *13*, 1595–1604.

(44) Holmes, A. A.; Umrigar, C. J.; Sharma, S. Excited states using semistochastic heat-bath configuration interaction. *J. Chem. Phys.* **2017**, *147*, 164111.

(45) Power, J. D.; Pitzer, R. M. Inequalities For Electron Repulsion Integrals. *Chem. Phys. Lett.* **1974**, *24*, 478–483.

(46) Helgaker, T.; Jørgensen, P.; Olsen, J. *Mol. Electron. Theory*; John Wiley & Sons, Ltd: Chichester, UK, 2014; pp 648–723.

(47) Overy, C.; Booth, G. H.; Blunt, N. S.; Shepherd, J. J.; Cleland, D.; Alavi, A. Unbiased reduced density matrices and electronic properties from full configuration interaction quantum Monte Carlo. *J. Chem. Phys.* **2014**, *141*, 244117.

(48) Umrigar, C. J.; Nightingale, M. P.; Runge, K. J. A diffusion Monte Carlo algorithm with very small timeâĂŘstep errors. *J. Chem. Phys.* **1993**, *99*, 2865–2890.

(49) Vigor, W. A.; Spencer, J. S.; Bearpark, M. J.; Thom, A. J. W. Minimising biases in full configuration interaction quantum Monte Carlo. *J. Chem. Phys.* **2015**, *142*, 104101.

(50) Flyvbjerg, H.; Petersen, H. G. Error estimates on averages of correlated data. *J. Chem. Phys.* **1989**, *91*, 461–466.

(51) Spencer, J. S.; Blunt, N. S.; Vigor, W. A.; Malone, F. D.; Foulkes, W. M. C.; Shepherd, J. J.; Thom, A. J. W. Open-Source Development Experiences in Scientific Software: The HANDE Quantum Monte Carlo Project. *J. Open Res. Softw.* **2015**, *3*, 1–6.

(52) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; Mcclain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-l. PYSCF : the Python-based simulations of chemistry framework. *WIREs Comput. Mol. Sci.* **2018**, *8:e1340*.

(53) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300–302.

(54) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.

(55) Thom, A. J. W.; Alavi, A. A combinatorial approach to the electron correlation problem. *J. Chem. Phys.* **2005**, *123*, 204106.

(56) Thom, A. J. W.; Alavi, A. Stochastic Perturbation Theory: A Low-Scaling Approach to Correlated Electronic Energies. *Phys. Rev. Lett.* **2007**, *99*, 143001.

(57) Kolodrubetz, M.; Clark, B. K. Partial node configuration-interaction Monte Carlo as applied to the Fermi polaron. *Phys. Rev. B* **2012**, *86*, 075109.

(58) Pederiva, F.; Roggero, A.; Schmidt, K. E. In *An Adv. Course Comput. Nucl. Phys. Bridg. Scales from Quarks to Neutron Stars*; Hjorth-Jensen, M., Lombardo, M. P., van Kolck, U., Eds.; Springer International Publishing: Cham, 2017; pp 401–476.

(59) Ohtsuka, Y.; Nagase, S. Projector Monte Carlo method based on configuration state functions. Test applications to the H4 system and dissociation of LiH. *Chem. Phys. Lett.* **2008**, *463*, 431–434.

(60) Walker, A. J. New fast method for generating discrete random numbers with arbitrary frequency distributions. *Electron. Lett.* **1974**, *10*, 127.

(61) Walker, A. J. An Efficient Method for Generating Discrete Random Variables with General Distributions. *ACM Trans. Math. Softw.* **1977**, *3*, 253–256.

(62) Kronmal, R. A.; Peterson, A. V. On the Alias Method for Generating Random Variables from a Discrete Distribution. *Am. Stat.* **1979**, *33*, 214–218.

(63) Knuth, D. E. *Art Comput. Program. / Vol. 2 , Semin. algorithms.*, 3rd ed.; Addison-Wesley: Reading, Mass. ; Harlow, pp 119–142.

(64) Schwarz, L. R. Projector Quantum Monte Carlo Methods for Linear and Nonlinear Wavefunction Ansatzes. Ph.D. thesis, Cambridge, 2017.

(65) Slater, J. C. The Theory of Complex Spectra. *Phys. Rev.* **1929**, *34*, 1293–1322.

(66) Condon, E. U. The Theory of Complex Spectra. *Phys. Rev.* **1930**, *36*, 1121–1133.

(67) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69–89.

(68) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(69) Vigor, W. A.; Spencer, J. S.; Bearpark, M. J.; Thom, A. J. W. Understanding and improving the efficiency of full configuration interaction quantum Monte Carlo. *J. Chem. Phys.* **2016**, *144*, 094110.

Graphical TOC Entry