Understanding and Improving the Efficiency of Full Configuration Interaction Quantum Monte Carlo

W. A. Vigor,[1] J. S. Spencer,[2,3] M. J. Bearpark,[1] and A. J. W. Thom[1,4]

[1] *Department of Chemistry, Imperial College London, Exhibition Road, London, SW7 2AZ, United Kingdom*

[2] *Department of Physics, Imperial College London, Exhibition Road, London, SW7 2AZ, United Kingdom*

[3] *Department of Materials, Imperial College London, Exhibition Road, London, SW7 2AZ, United Kingdom*

[4] *University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, United Kingdom*

Within Full Configuration Interaction Quantum Monte Carlo, we investigate how the statistical error behaves as a function of the parameters which control the stochastic sampling. We define the inefficiency as a measure of the statistical error per particle sampling the space and per timestep and show there is a sizeable parameter regime where this is minimised. We find that this inefficiency increases sublinearly with Hilbert space size and can be reduced by localising the canonical Hartree–Fock molecular orbitals, suggesting that the choice of basis impacts the method beyond that of the sign problem.

I. INTRODUCTION

The FCIQMC method[1,2] is a stochastic algorithm that has enabled calculation of the ground state energy of the largest molecules to date[3] to FCI accuracy, i.e. exact given a fixed basis set. It has also found success calculating the ground state energy of model Hamiltonians such as the uniform electron gas[4] and the three band Hubbard model[5], as well as solid state systems.[6] Recent developments mean that properties beyond the ground state energy, such as forces[7] and excited state energies[8–10] can also be calculated. Further, efficient stochastic implementations of other methods such as second order Møller Plesset Perturbation theory (MP2)[11–13], Density Functional Theory (DFT)[14] and Coupled Cluster theory[15] have been developed, forming a new field of stochastic computational chemistry. These Monte Carlo methods differ from the more traditional Diffusion Monte Carlo method[16], which requires the use of a fixed node approximation to prevent collapse of the wavefunction to a bosonic state. The eventual hope is that these methods will allow the properties and reactivity of large molecules to be investigated to unprecedented accuracy.

Quantum Monte Carlo simulations are plagued by the fermion sign problem: the quantity being sampled can be either positive or negative which may cause the calculation to converge to the wrong answer.[17,18] In FCIQMC the sign problem appears as the critical population of particles required to correctly sample the ground state wavefunction.[18] Whilst the scaling of this with system size has been investigated in a number of systems[18–20], the effect of it on the stochastic error has not been well studied. The (systematically improvable) initiator approximation[2] can reduce the critical population by many orders of magnitude. However, the impact of this key advance on the stochastic error has similarly not been thoroughly investigated.

The efficiency of a stochastic algorithm can be mea-sured by computational cost (CPU time and memory requirements) necessary for convergence to a given stochastic error. The FCIQMC algorithm is not a 'black box' method and the manner in which sampling is performed can impact the computational resources used whilst performing a calculation and the resultant statistical error bar. We believe that understanding the efficiency of a FCIQMC calculation is important for two main reasons. Firstly so that one can use a computational budget as effectively as possible, ideally leading to the ability to perform FCIQMC calculations automatically and efficiently without careful tuning of input parameters. Secondly it helps compare the effectiveness of different algorithms based upon and additional approximations to FCIQMC.

Previous studies of the efficiency of FCIQMC investigated the standard error as a function of computer time and focused on comparing very different algorithms.[21,22] Here we take a different tack, running many empirical calculations to understand how the stochastic sampling of the wavefunction affects the error bar for FCIQMC in general. Once we understand this effect, we form a metric which is independent of the parameter choices and use it to measure how the error bar scales with system size.

We begin with an overview of the FCIQMC method. We then explore the dependence of the statistical error bar on the total population and the timestep in Section IV and on the system size, both basis set and number of electrons, in Section V. Finally, we discuss in Section VI how the total number of particles (relative to the plateau height) affects the stochastic error bar and draw some conclusions about how to choose parameters to maximise the efficiency of the FCIQMC algorithm 'a priori' from the plateau height.

II. A RECAP OF THE FCIQMC METHOD

The FCIQMC algorithm[1,2] can be viewed as a stochastic power method. On every iteration the representation of the wavefunction, $\psi(\tau)$, at imaginary time τ is updated by sampling the action of the operator:

$$\psi(\tau + \delta\tau) = \left(1 - (\hat{H} - S)\delta\tau\right)\psi(\tau) \qquad (1)$$

where $\hat{H}$ is the Hamiltonian, S is an offset used to control normalisation, and $\delta\tau$ is the timestep. As long as $\psi(0)$ has a non-zero overlap with it, the ground state will be all that remains once τ is large enough, assuming $\delta\tau$ is sufficiently small[18] and S is carefully controlled. $\langle D_i|\psi(\tau)\rangle$ is represented by a number of (signed) unit weights located on $|D_i\rangle$; we call a single unit a psi-particle or psip.[23] Other choices of discretisation are possible[16]; for the purposes of this work we consider the simplest case.

Eq. (1) is sampled in three stages to make up a single timestep of $\delta\tau$[1]:

Spawning: each psip (with weight w_i) attempts to spawn a child psip on a randomly selected $|D_j\rangle$ with probability $|\langle D_i|\hat{H}|D_j\rangle|\delta\tau/p(j|i)$, where $p(j|i)$ is the probability that $|D_j\rangle$ is selected given the parent psip is on $|D_i\rangle$ and with sign $\mathrm{sgn}(-\langle D_i|\hat{H}|D_j\rangle w_i)$.

Death: each psip dies (is removed from the simulation) with probability $|K|\delta\tau$, where $K = \langle D_i|\hat{H}|D_i\rangle - S$, if $K < 0$; for $K > 0$, instead, a copy of the psip is made with probability $K\delta\tau$.

Annihilation: psips on the same determinant with opposite signs cancel.

Initially S is set to the Hartree–Fock energy, which is larger (less negative) than the FCI energy and this causes the population $N_p(\tau)$ (total number of psips) to begin to grow exponentially. If there is a sign problem[24] $N_p(\tau)$ will plateau as the sign structure is projected out. It is only after this point that $\psi(\tau)$ becomes a stochastic representation of the wavefunction and $N_p(\tau)$ grows exponentially again.[18] To counter this exponential growth S is periodically adjusted every A timesteps according to:

$$S(\tau + A\delta\tau) = S(0) - \xi \log \frac{N_p(\tau + A\delta\tau)}{N_s}, \qquad (2)$$

where $S(0)$ is the initial value of the shift, N_s is the population at the end of the equilibration phase and ξ, the damping factor, is usually fixed during a simulation.[25]

In addition to the shift, the energy of the system can be estimated via the projected estimator, which typically has a smaller statistical noise:

$$E_{\mathrm{Proj}} = \frac{\langle D_0|\hat{H}e^{-\hat{H}\tau}|D_0\rangle}{\langle D_0|e^{-\hat{H}\tau}|D_0\rangle} = \frac{\langle \sum_{i\neq 0} H_{0i}n_i\rangle_\tau}{\langle n_0\rangle_\tau}, \qquad (3)$$

where $n_i(\tau)$ is the (signed) number of psips on determinant i in $|\psi(\tau)\rangle$ and $\langle\ldots\rangle_\tau$ represents the time average.

The initiator approximation[2] can dramatically reduce the sign problem in FCIQMC.[6,7,9,20,22,26–28] In initiator FCIQMC (iFCIQMC), previously unoccupied determinants can only be spawned onto from determinants with a population above a certain threshold; a threshold of 3 is typical and used throughout this paper. The initiator approximation introduces a systematic error in the sampling of Eq. (1), which is evident in both estimates of the energy, with unbiased sampling restored in the limit of an infinite population of psips. A monotonic convergence to the FCI energy as a function of $\langle N_p(\tau)\rangle$ has been observed for many simple systems,[2] though this is not universally the case.

III. SYSTEMS STUDIED

In this paper we study the following systems:

1. An isolated neon atom in the following Dunning basis sets[29]: cc-pVDZ, aug-cc-pVDZ, cc-pVTZ, aug-cc-pVTZ, cc-pVQZ, aug-cc-pVQZ.

2. The hydrogen fluoride, HF, molecule in a cc-pVDZ basis at bond lengths of $R = R_0$, $R = 1.5R_0$ and $R = 2R_0$, where $R_0 = 0.91622$ Å, the Hartree–Fock equilibrium bond length in this basis.

3. Chains of between 2 and 7 helium atoms at intervals of 3 Å in a 6-31G basis set.[30]

4. Chains of between 5 and 7 helium atoms at intervals of 3 Å in a 6-31G basis set with localised molecular orbitals. We used Pipek–Mezey localisation[31] to separately localise the occupied and unoccupied Hartree–Fock molecular orbitals.

Hartree–Fock calculations were performed using Q-Chem[32], with local modifications to obtain the required integrals. FCIQMC calculations were performed and analysed using HANDE[33] and figures plotted using matplotlib.[34] Unless otherwise stated, atomic units are used throughout.

IV. POPULATION DYNAMICS AND THE STOCHASTIC ERROR

If the Hamiltonian doesn't have a 'sign problem' in FCIQMC, it is impossible for psips of opposite signs to be generated on the same determinant and so annihilation cannot occur. In this case each psip samples the wavefunction independently from each other and hence the error in the estimate of the energy, σ_E, must behave as:

$$\sigma_E = \frac{a}{\sqrt{\langle N_p\rangle N\delta\tau}}, \qquad (4)$$

where a is a constant of proportionality and N is the number of timesteps from which σ_E is estimated. Eq. (4) arises because the contribution to the estimate of E from each psip can be combined in a similar way to the contribution for running from multiple timesteps. Increasing $\delta\tau$ simply increases the probability of spawning/death and so decreases σ_E in the same way as increasing N or $\langle N_p \rangle$.

a can be estimated from a FCIQMC calculation using σ_E estimated by blocking analysis of the projected energy and $\langle N_p \rangle$. For consistency throughout this paper, we use an automatic approach to calculate the optimal block length[35], which has previously been applied to DMC[36] and is implemented in pyblock.[37]

For systems with a 'sign problem', which is the usual case, a will depend on $\langle N_p \rangle$ and $\delta\tau$. We will call $a(\langle N_p \rangle, \delta\tau)$ the inefficiency and note from Eq. (4) that a smaller a implies a more efficient calculation. In Secs. IV A and IV B we investigate the dependence of a on $\langle N_p \rangle$ and $\delta\tau$ respectively.

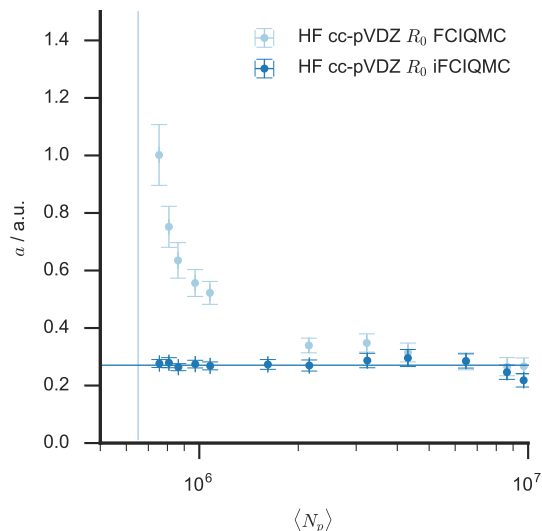## A. Effect of population on inefficiency of FCIQMC simulations

FIG. 1. The inefficiency a as a function of the average number of psips for HF in the cc-pVDZ basis, $R_0 = 0.91622\,\text{Å}$, and $\delta\tau = 0.00175$ for both FCIQMC and iFCIQMC. The plateau height is $\sim 6.5 \times 10^5$ (measured using the method described in 38) and is indicated by the vertical line. The horizontal line shows a fit to the iFCIQMC data for constant a. The initiator error is converged below the stochastic error in all iFCIQMC calculations shown; a may not remain constant where this is not true. For a population of $\langle N_p \rangle = 50000$ in iFCIQMC (not shown), we find $a = 0.318(18)$ compared to $a = 0.2682(34)$ from the fit, showing that sufficiently small populations do have an effect on the inefficiency in iFCIQMC calculations.

The behaviour of a with $\langle N_p \rangle$ in FCIQMC simulations is shown in Fig. 1 for hydrogen fluoride. This behaviour has been seen in a wide range of systems, which are included in the Supplemental Information.[39] Owing to the sign problem, only after the population reaches the plateau does the vector of psips become a stochastic representation of the eigenvector.[18] Populations below the plateau could either have a divergent σ_E or an incorrect average correlation energy with a finite σ_E; we see the former behaviour. For populations greater than the plateau, a decays as a function of $\langle N_p \rangle$ and tends to a finite constant in the large population limit.

If the initiator approximation is used, then we find that a is a constant as a function of $\langle N_p \rangle$ for populations much smaller than the plateau and, for a fixed $\delta\tau$, the same as a in FCIQMC in the large $\langle N_p \rangle$ limit. We call this limit on a the iFCIQMC limit. For example in Fig. 1 inefficiency in FCIQMC doesn't hit the iFCIQMC limit until about $\langle N_p \rangle = 2 \times 10^6$, approximately 3 times the plateau height. This is an important point: if the largest population affordable in FCIQMC is not sufficient to reach the iFCIQMC limit (but is sufficient to exceed the plateau), then the initiator approximation is still useful as it provides a significant reduction in the stochastic error for the same computational cost. It may, however, be difficult to quantify if the introduction of an initiator error is a price worth paying for a potentially significant reduction in statistical error bar.

This behaviour has implications about how best to run parallel FCIQMC and iFCIQMC simulations given a fixed amount of computational resources.

In the canonical parallel FCIQMC implementation[40] the Hilbert space is partitioned over the processors, resulting in efficient distribution of the memory demands across the processors. Psips spawned from a parent determinant in one part of the Hilbert space onto a child determinant in another part of the space are communicated between processors. It is this communication overhead that limits the parallel scalability of FCIQMC.

The simplest way to parallelise a Monte-Carlo algorithm is to run independent calculations and combine statistics gathered in each calculation. The overall stochastic error scales as $1/\sqrt{N_I}$, where N_I is the number of independent calculations. Therefore once the population is such that a reaches the iFCIQMC limit, it is more efficient to use multiple independent simulations to minimise interprocess communication.

B. Effect of timestep on inefficiency of FCIQMC simulations

a is approximately constant for a sufficiently small timestep, $\delta\tau$, and otherwise a increases with $\delta\tau$; the value of $\delta\tau$ after which a is non-constant is system dependent. We have seen this behaviour for all systems investigated in this paper (see Supplemental Information[39]) and for FCIQMC and iFCIQMC. Fig. 2 shows this for an isolated
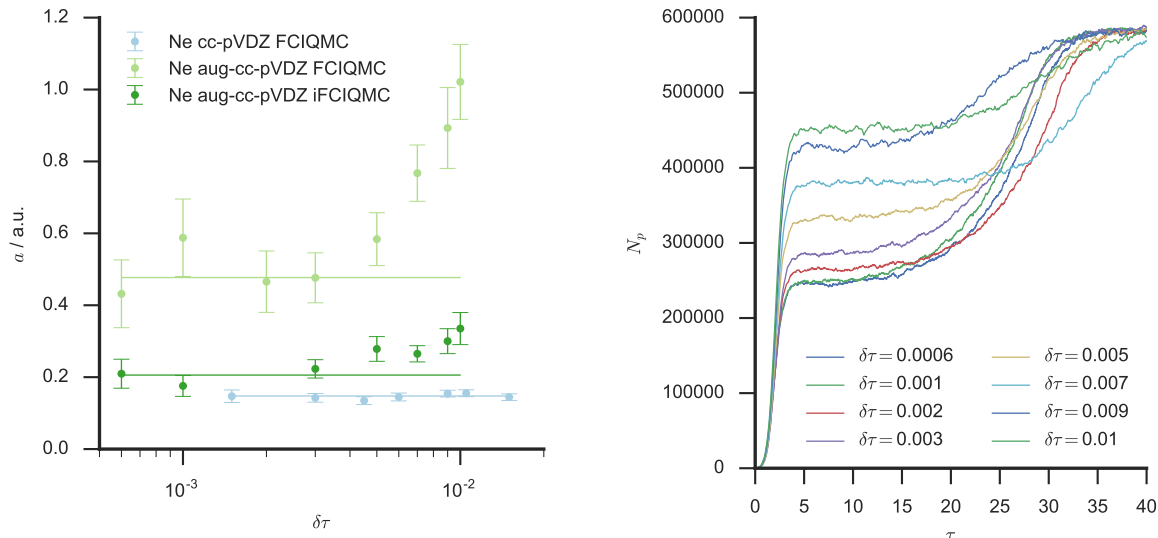
FIG. 2. Left: The inefficiency a as a function of $\delta\tau$ for Ne in cc-pVDZ ($\langle N_p \rangle \approx 46100$) and aug-cc-pVDZ bases ($\langle N_p \rangle \approx 590000$). Once $\delta\tau$ is large enough to raise the plateau height above $\langle N_p \rangle$, a diverges in a similar way as in Section IV A. Right: The plateau height as a function of the timestep for FCIQMC on Ne in aug-cc-pVDZ; note the region in the plateau is roughly minimum matches the similar region for a.

neon atom in cc-pVDZ and aug-cc-pVDZ bases.

A previous investigation[41] has shown that the plateau height exhibits a similar behaviour to a as a function of $\delta\tau$ in FCIQMC calculations. As shown in Fig. 2 for the Ne atom (and in the Supplemental Information[39] for other systems) the plateau height appears to be a good metric for when a remains constant as a function of $\delta\tau$. This is useful as the plateau height is easier and cheaper to measure than statistical accumulation of the energy and hence evaluation of a. We denote the largest $\delta\tau$ such that a is constant as $\delta\tau_0$.

In the Supplemental Information[39] we investigate whether $\langle N_p \rangle$ has an impact on $\delta\tau_0$ and find this is not the case. We do however find that a increases faster as a function of $\delta\tau$ when $\delta\tau > \delta\tau_0$ for a small $\langle N_p \rangle$ than a large $\langle N_p \rangle$ and hence the most efficient timestep has an implicit dependency on $\langle N_p \rangle$.

The population dynamics can become unstable if the $\delta\tau$ is set large enough such that the exponential growth cannot be countered by population control. We find that the point at which the population explodes is beyond $\delta\tau_0$ for the systems studied here with the exception of Ne cc-pVDZ (see Fig. 2 left) which doesn't have a noticeable plateau.[42]

V. THE SCALING OF INEFFICIENCY AS A FUNCTION OF SYSTEM SIZE

In the previous sections we investigated both the scaling of the inefficiency, a, as a function of (mean) population, $\langle N_p \rangle$, and timestep, $\delta\tau$. Following this, we can define a metric that lets us investigate the scaling of the error bar as we change system. The simplest such metric is to find the minimum value of a, $a_{\min}$, i.e. when it is a constant as a function of $\langle N_p \rangle$ and $\delta\tau$. This requires calculations with $\delta\tau < \delta\tau_0$ and either the initiator approximation (with the initiator error converged) or a large $\langle N_p \rangle$. The behaviour of $a_{\min}$ as a function of Hilbert space size is shown in Fig. 3 for all systems studied in this work: the neon atom, hydrogen fluoride at different bond lengths and chains of helium atoms in the canonical and localised basis sets. We see a sublinear relationship between the scaling of $a_{\min}$ with size of Hilbert space for the same chemical species and a significant impact when the degree of strong correlation is increased.

FCIQMC is very efficient at finding the ground state of the neon atom: f_c, the ratio of plateau height to Hilbert space size, is $\sim 10^{-4}$.[1] The size of the Hilbert space scales factorially with the basis set yet there is a sublinear scaling of the stochastic error with the Hilbert space size.

In contrast f_c scales superlinearly for chains of helium atoms separated by 3 Å in the canonical Hartree–Fock basis: for He_4 $f_c = 0.3$ whereas for He_7 $f_c = 0.97$. However, $a_{\min}$ again seems to scale sublinearly with the size of the Hilbert space. The similar behaviour, within statistical error bars, of $a_{\min}$ for Ne atom and helium chains suggests sublinear scaling of the stochastic error in FCIQMC with Hilbert space size for weakly correlated systems with no strong dependence on the FCIQMC sign problem. In a similar style study in DMC[43], Nemec et al. investigated computational scaling as a function of the number of hydrogen atoms at a large separation, finding that the computer time required to achieve a fixed error bar scales
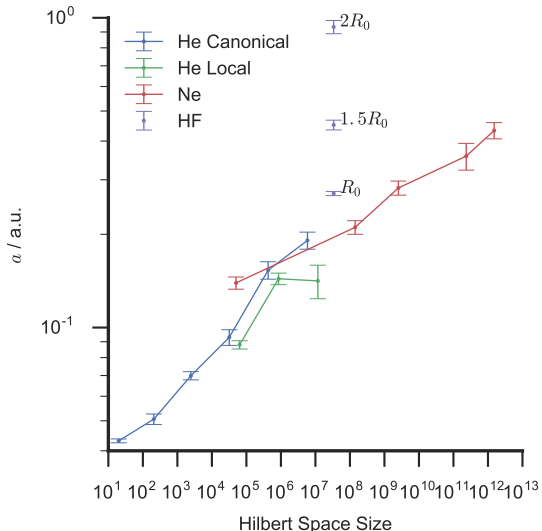
FIG. 3. The scaling of $a_{\min}$ as a function of system size. Section III describes the systems studied in more detail. $a_{\min}$ was calculated by fitting a for iFCIQMC calculations with differing populations with the exception of Ne cc-pVDZ, He2, He3 and He4, which have no noticeable plateau, where many FCIQMC calculations were used. The timestep was set to be sufficiently small such that a is constant. The Supplemental Information[39] shows the studies used to find $a_{\min}$ for each system.

exponentially as the square root of the system size. As the Hilbert space size scales loosely exponentially with the system size, we see this trend is similar, though with such few systems studied it seems premature to generalize this trend.

Localisation of the molecular oribitals breaks symmetry and so increases the size of the accessible Hilbert space. However we find in this case that the plateau height decreases[44]; for example localisation causes the plateau for He7 to decrease from 5.7×10^6 ($f_c = 0.97$) to 4.6×10^6 ($f_c = 0.39$). Importantly localisation also appears to decrease $a_{\min}$ compared to using the canonical orbitals, especially for He7, though it is difficult to draw definitive conclusions from three data points with non-negligible standard errors.

The behaviour of $a_{\min}$ is not solely governed by the size of the Hilbert space, however. Stretching hydrogen fluoride in a fixed basis set (and hence Hilbert space) causes the correlation energy to increase. We find that $a_{\min}$ increases significantly with bond length but have not found a simple linear scaling with the correlation energy or plateau height. Given this behaviour doesn't fit with the weakly correlated systems studied above, investigation of the scaling of the error bar in FCIQMC in the strongly correlated limit would be an interesting topic for future investigation.

VI. CONCLUSIONS

We have defined a metric, *inefficiency*, for measuring the stochastic efficiency of FCIQMC calculations and investigated its behaviour as a function of the parameters controlling the population dynamics and as a function of system size. We have found that a sizeable reduction in the stochastic error is possible by increasing the population of psips and using the largest possible timestep such that the plateau height remains constant.[45] The optimal timestep is transferable between FCIQMC and iFCIQMC simulations. The efficiency decreases (and inefficiency increases) sublinearly with the size of the Hilbert space in weakly correlated systems. The population required to exceed the plateau height or converge the initiator approximation becomes a limiting factor much faster than the rise in the stochastic error bar for a given computational effort. This suggests that if improved approximations which reduce the sign problem can be made, then the stochastic noise will not be insurmountable when treating yet larger systems.

This analysis ignores the impact of population and timestep on the computational cost of the calculation, which is detailed in Appendix A. Ideally we would like to set the population and timestep such that the stochastic error decays as fast as possible as a function of the computational effort. This analysis doesn't change the guidelines given above: i) the population should be large enough such that the inefficiency metric, a, is converged to the iFCIQMC limit; ii) the largest timestep for which the plateau remains constant represents a lower bound on the most efficient timestep and is difficult to improve upon without running lots of expensive calculations at different timesteps.

Orbital localisation was found to be effective for reducing the plateau height, and thus the cost of a FCIQMC calculation for chains of helium atoms, as well improving the stochastic efficiency. Consequently, it may be worthwhile investigating FCIQMC algorithms tuned to localised orbitals.

More broadly, we have provided a framework for assessing the statistical efficiency of FCIQMC and related methods. Given the recent activity in such methodological development[15,21,46–49], this approach offers a useful means to fairly compare alternative implementations and algorithms.

ciety for a University Research Fellowship under Grant No. UF110161. Calculations were performed using the Imperial College High Performance Computing Service.[50]

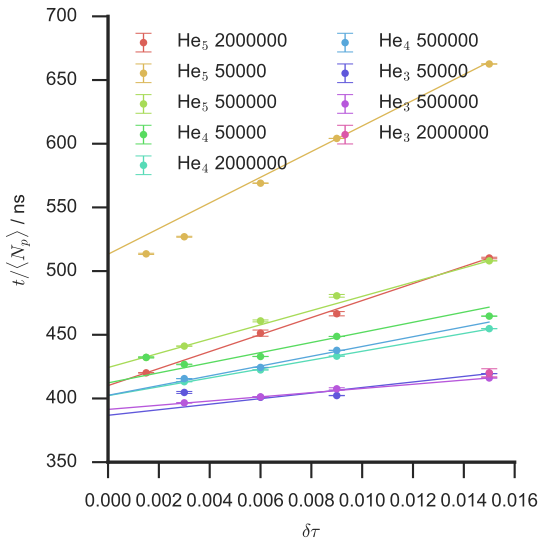## Appendix A: Computational cost of an FCIQMC calculation

FIG. 4. The computer time of a single iteration per psip as a function of the timestep. There is a linear relationship (Eq. (A1)) with the timestep and a near-linear relationship with the number of psips for a large population. All calculations were run on a single core of an Intel Core i7-2600 processor.

An efficient FCIQMC implementation uses a sparse storage scheme to store the list of psips: a representation of a determinant along with the number of psips occupying the determinant are stored together.[40] The expensive steps are evolving the psips (spawning and death) and annihilation (combining psips on the same determinant).

The cost of spawning and death is linear in the population; the death step can be efficiently performed on a per-determinant rather than per-psip basis but the effect of this can absorbed into a system-dependent prefactor. The prefactor also includes the cost of selecting a random excitation and calculating the necessary Hamiltonian matrix elements.

The cost of annihilation depends on the number of psips created in a single timestep. Assuming the calculation has equilibrated at a population in excess of the critical population, the number of psips created scales linearly with the population and the timestep with a rate dependent on the average absolute value of the off-diagonal elements in the Hamiltonian matrix. Whilst annihilation can be performed with a linear scaling[40], the implementation in HANDE currently uses a binary search approach

that introduces a logarithmic dependency on the number of occupied determinants. For the purposes of this analysis, we shall neglect the logarithmic dependency and instead focus on the optimal case.

Under these assumptions the average total amount of computer time to perform a single timestep $\langle t \rangle$ is:

$$\langle t \rangle = C_1 \langle N_p \rangle + C_2 \delta\tau \langle N_p \rangle. \tag{A1}$$

C_1 and C_2 depend on both the computer architecture and the chemical system of interest. All timing calculations were run on one processor (core) of the same machine.[51]

In Fig. 4 we show the computational time, t, of a single timestep per psip for a range of different populations and systems. If Eq. (A1) was obeyed exactly then there to be no dependence between the number of psips and the intercept or gradient of Fig. 4 for a given system; the approximations made are good at large $\langle N_p \rangle$. The approximation is less suitable for a small population as it assumes that the population is such that the average number of occupied determinants is approximately constant.

From Eq. (A1) and Eq. (4), it follows that

$$\sigma_E = \frac{a(\langle N_p \rangle, \delta\tau)\sqrt{C_1/\delta\tau + C_2}}{\sqrt{N\langle t \rangle}}. \tag{A2}$$

As $N\langle t \rangle$ is the total computer time for running for N steps, minimising $a' = a\sqrt{C_1/\delta\tau + C_2}$ gives the most efficient population dynamics. Whilst the exact form of $a(\langle N_p \rangle, \delta\tau)$ is not known, we can draw qualitative conclusions from the behaviour observed in IV A and IV B: i) a' and a share the same dependence on $\langle N_p \rangle$; FCIQMC calculations should therefore be performed with sufficiently large population to place it in the iFCIQMC limit; ii) if $\delta\tau < \delta\tau_0$ then a' will decrease as $\delta\tau^{-1/2}$, given that a is constant with $\delta\tau$ in this region. If $\delta\tau > \delta\tau_0$ a will increases with $\delta\tau$ which will compete with (and dominate) the $\sqrt{C_1/\delta\tau + C_2}$ term, making it difficult to quantify what will happen to a'. Thus $\delta\tau_0$ forms a lower bound on the most efficient timestep.

[1]G. H. Booth, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **131**, 054106 (2009).
[2]D. Cleland, G. H. Booth, and A. Alavi, J. Chem. Phys. **132**, 41103 (2010).
[3]C. Daday, S. Smart, G. H. Booth, A. Alavi, and C. Filippi, J. Chem. Theory Comput. **8**, 4441–4451 (2012).
[4]J. J. Shepherd, G. H. Booth, and A. Alavi, J. Chem. Phys. **136**, 244101 (2012).
[5]L. R. Schwarz, G. H. Booth, and A. Alavi, Phys. Rev. B **91**, 045139 (2015).
[6]G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, Nature **493**, 365–370 (2013).
[7]R. E. Thomas, D. Opalka, C. Overy, P. J. Knowles, A. Alavi, and G. H. Booth, J. Chem. Phys. **143**, 054108 (2015).
[8]G. H. Booth and G. K.-L. Chan, J. Chem. Phys. **137**, 191102 (2012).
[9]N. S. Blunt, A. Alavi, and G. H. Booth, Phys. Rev. Lett. **115**, 050603 (2015).
[10]N. S. Blunt, S. D. Smart, G. H. Booth, and A. Alavi, ArXiv e-prints (2015), arXiv:1508.04680 [physics.comp-ph].

[11] A. J. W. Thom and A. Alavi, Phys. Rev. Lett. **99**, 143001 (2007).

[12] S. Y. Willow, K. S. Kim, and S. Hirata, J. Chem. Phys. **137**, 204122 (2012).

[13] S. Y. Willow, M. R. Hermes, K. S. Kim, and S. Hirata, J. Chem. Theory Comput. **9**, 4396–4402 (2013).

[14] R. Baer, D. Neuhauser, and E. Rabani, Phys. Rev. Lett. **111**, 106402 (2013).

[15] A. J. W. Thom, Phys. Rev. Lett. **105**, 263004 (2010).

[16] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, J. Chem. Phys. **99**, 2865–2890 (1993).

[17] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566–569 (1980).

[18] J. S. Spencer, N. S. Blunt, and W. M. C. Foulkes, J. Chem. Phys. **136**, 054110 (2012).

[19] D. M. Cleland, G. H. Booth, and A. Alavi, J. Chem. Phys. **134**, 024112 (2011).

[20] J. J. Shepherd, G. E. Scuseria, and J. S. Spencer, Phys. Rev. B **90**, 155130 (2014).

[21] F. R. Petruzielo, A. A. Holmes, H. J. Changlani, M. P. Nightingale, and C. J. Umrigar, Phys. Rev. Lett. **109**, 230201 (2012).

[22] N. S. Blunt, S. D. Smart, J. A. F. Kersten, J. S. Spencer, G. H. Booth, and A. Alavi, J. Chem. Phys. **142**, 184107 (2015).

[23] J. B. Anderson, J. Chem. Phys. **63**, 1499–1503 (1975).

[24] There is a sign problem in FCIQMC if it is possible that psips of opposite signs can be generated on the same determinant.[18].

[25] As discussed in Ref. 52 this is equivalent to the original formulation of the shift in Ref. 1.

[26] G. H. Booth, D. Cleland, A. Alavi, and D. P. Tew, J. Chem. Phys. **137**, 1–11 (2012).

[27] C. Overy, G. H. Booth, N. S. Blunt, J. J. Shepherd, D. Cleland, and A. Alavi, J. Chem. Phys. **141**, 1–11 (2014).

[28] J. J. Shepherd, G. Booth, A. Grüneis, and A. Alavi, Phys. Rev. B **85**, 081103 (2012).

[29] T. H. Dunning, J. Chem. Phys. **90**, 1007–1023 (1989).

[30] R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **54**, 724–728 (1971).

[31] J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916–4926 (1989).

[32] Y. e. a. Shao, Mol. Phys. **113**, 184–215 (2015).

[33] http://www.hande.org.uk/.

[34] J. D. Hunter, Comput. Sci. Eng. **9**, 90–95 (2007).

[35] U. Wolff, Comput. Phys. Commun. **156**, 143 – 153 (2004).

[36] R. M. Lee, G. J. Conduit, N. Nemec, P. Lopez Rios, and N. D. Drummond, Phys. Rev. E **83**, 066706 (2011).

[37] See https://github.com/jsspencer/pyblock for code.

[38] J. S. Spencer and A. J. W. Thom, ArXiv e-prints (2015), arXiv:1511.05752 [physics.chem-ph].

[39] W. A. Vigor, J. S. Spencer, M. J. Bearpark, and A. J. W. Thom, "Supplemental material for "understanding the efficiency of full configuration interaction quantum monte carlo"," .

[40] G. H. Booth, S. D. Smart, and A. Alavi, Mol. Phys. **112**, 1855–1869 (2014).

[41] G. Booth, *A novel quantum Monte Carlo method for molecular systems*, Ph.D. thesis, University of Cambridge (2010).

[42] When running the calculations for Fig. 2 top right we found that the population explodes when $\delta\tau$ is slightly greater than 0.05 for Ne cc-pVDZ.

[43] N. Nemec, Phys. Rev. B **81**, 035119 (2010).

[44] Note that this behaviour is system specific. Ref. 41 found that systems where the plateau height was unaffected by orbital localisation.

[45] This can be easily assessed prior to production calculations given that propagation to the plateau is computationally cheap compared to an entire FCIQMC calculation and the plateau height can easily be found in an automated fashion.[20,38].

[46] N. S. Blunt, T. W. Rogers, J. S. Spencer, and W. M. C. Foulkes, Phys. Rev. B **89**, 245124 (2014).

[47] F. D. Malone, N. S. Blunt, J. J. Shepherd, D. K. K. Lee, J. S. Spencer, and W. M. C. Foulkes, J. Chem. Phys. **143**, 044116 (2015).

[48] S. Ten-no, J. Chem. Phys. **138**, 164126 (2013).

[49] M. Kolodrubetz and B. K. Clark, Phys. Rev. B **86**, 075109 (2012).

[50] http://www.imperial.ac.uk/ict/services/teachingandresearchservices/highperformancecomputing.

[51] By running on a single core we avoid complications resulting from N_p affecting load balancing, which will in turn affect t. Load balancing has been extensively investigated in Ref. 40.

[52] W. A. Vigor, J. S. Spencer, M. J. Bearpark, and A. J. W. Thom, J. Chem. Phys. **142**, 104101 (2015).