Communications: Survival of the fittest: Accelerating convergence in full configuration-interaction quantum Monte Carlo

Deidre Cleland, George H. Booth and Ali Alavi

View Online   Export Citation

ARTICLES YOU MAY BE INTERESTED IN

Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space
The Journal of Chemical Physics **131**, 054106 (2009); https://doi.org/10.1063/1.3193710

The sign problem and population dynamics in the full configuration interaction quantum Monte Carlo method
The Journal of Chemical Physics **136**, 054110 (2012); https://doi.org/10.1063/1.3681396

NECI: N-Electron Configuration Interaction with an emphasis on state-of-the-art stochastic methods
The Journal of Chemical Physics **153**, 034107 (2020); https://doi.org/10.1063/5.0005754

Communications: Survival of the fittest: Accelerating convergence in full configuration-interaction quantum Monte Carlo

Deidre Cleland, George H. Booth, and Ali Alavi[a)]
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

We provide a very simple adaptation of our recently published quantum Monte Carlo algorithm in full configuration-interaction (Slater determinant) spaces which dramatically reduces the number of walkers required to achieve convergence. A survival criterion is imposed for newly spawned walkers. We define a set of initiator determinants such that progeny of walkers spawned from such determinants onto unoccupied determinants are able to survive, while the progeny of walkers not in this set can survive only if they are spawned onto determinants which are already occupied. The set of initiators is originally defined to be all determinants constructable from a subset of orbitals, in analogy with complete-active spaces. This set is dynamically updated so that if a noninitiator determinant reaches an occupation larger than a preset limit, it becomes an initiator. The new algorithm allows sign-coherent sampling of the FCI space to be achieved with relatively few walkers. Using the N_2 molecule as an illustration, we show that rather small initiator spaces and numbers of walkers can converge with submilliHartree accuracy to the known full configuration-interaction (FCI) energy (in the cc-pVDZ basis), in both the equilibrium geometry and the multiconfigurational stretched case. We use the same method to compute the energy with cc-pVTZ and cc-pVQZ basis sets, the latter having an FCI space of over 10^{15} with very modest computational resources. © *2010 American Institute of Physics*. [doi:10.1063/1.3302277]

We have recently published an exact algorithm for simulation of Fermion systems in Slater determinant spaces based on stochastic population dynamics of a set of annihilating positive and negative walkers, which simulates the underlying imaginary time Schrödinger equation of the interacting Hamiltonian.[1] We demonstrated that the algorithm converges onto the full configuration-interaction (FCI) energy and wave function, as long as the number of walkers in the simulation exceeds a (system-dependent) critical number (N_c). N_c was generally found to scale with the size of the FCI space (N_{det}), although for some systems the fraction $f_c = N_c/N_{det}$ could be as small as $f_c \approx 10^{-5}$ whereas for other systems it was close to 1, meaning that we needed almost as many walkers as there are determinants to achieve convergence.

Discovering methods which reduce the number of walkers required to achieve convergence has been the source of much investigation in our group. In this communication we will describe a very simple strategy which achieves this in a dramatic way. For example, we have been able to reduce the number of walkers needed to converge N_2 cc-pVDZ (with an FCI space of 5.4×10^8 determinants) from 2.7×10^8 to 10^5 while achieving a sub-mE_h level of accuracy. Translated into computer resources, this implies a gain of three orders of magnitude. Furthermore, we show that similar numbers of walkers can be used to converge the vastly larger problem of N_2 with a cc-pVQZ basis which has an FCI space exceeding 10^{15}.

We briefly outline the original algorithm; more details being provided in Ref. 1. We assume that we have at our disposal a set of $2M$ orthonormal one-particle orbitals [usually, but not necessarily, restricted Hartree–Fock (HF) orbitals] and the associated one- and two-body molecular integrals (which have been generated by the HF solver). The latter are used to compute matrix elements of the full Hamiltonian operator according to standard prescriptions. The basic entity in our simulation is a signed walker which permanently lives on the Slater determinant it was born. At each cycle of the algorithm, a walker on determinant $D_\mathbf{i}$ attempts to die with a probability given by $p_d = \tau(\langle D_\mathbf{i}|H - E_0|D_\mathbf{i}\rangle - S)$ where S is an energy shift and τ the timestep. In addition to this death/cloning step, each walker attempts to spawn a new walker at a connected determinant $D_\mathbf{j}$ with a probability $p_s = [\tau|\langle D_\mathbf{j}|H|D_\mathbf{i}\rangle|]/[p_{gen}(\mathbf{j}|\mathbf{i})]$, where $p_{gen}(\mathbf{j}|\mathbf{i})$ is the generation probability of $\mathbf{j}$ from $\mathbf{i}$. The sign of the child spawned is determined by the sign of the Hamiltonian matrix element and the parent; it has the same sign as the parent if the matrix element is negative and opposite to the parent if the matrix element is positive. After each walker has attempted the two steps above, there is a global annihilation step in which walkers of the opposite sign on the same determinant are removed from the simulation. The simulation can be run in constant shift (S) mode or in constant of number of walkers (N_w) mode. In the latter case, S is adjusted regularly in order to keep N_w approximately fixed. As discussed in Ref. 1 in a converged calculation, the time-averaged value of S equals the correlation energy. The total energy can also be measured directly from the long-time ensemble average of the walker distribution, according to the projection formula

---

a)Electronic mail: asa10@cam.ac.uk.

$$\langle E \rangle = E_0 + \sum_{\mathbf{j}} \langle D_{\mathbf{j}}|H|D_0\rangle \frac{\langle N_{\mathbf{j}}\rangle}{\langle N_0\rangle}, \qquad (1)$$

where $\mathbf{j}$ runs over the single and double excitations of a reference determinant D_0 (here taken to be the HF determinant) and $N_{\mathbf{j}}$ being the population of walkers on determinant $\mathbf{j}$. $E_0 = \langle D_0|H|D_0\rangle$ is the energy of the reference determinant D_0.[2]

We now introduce an additional rule in the algorithm, *which defines a survival criterion for newly spawned walkers*: some determinants (dubbed *initiator* determinants by analogy to chain-reactions) are endowed with the ability to spawn progeny onto unoccupied determinants. However, progeny of the remaining noninitiator determinants survive only if they are spawned onto a determinant which is already occupied. The exception to this rule is the special case whereby two noninitiator determinants spawn walkers of the same sign onto a third previously unoccupied determinant in one iteration. This is deemed to be a sign-coherent event and thus simultaneously spawned walkers such as these are allowed to live. The set of initiator determinants is chosen in two ways. First, a fixed set of initiator determinants are chosen according to a complete-active-space (CAS) criterion; i.e., a set (n, m) of n electrons distributed in m spatial orbitals about the Fermi energy, such that a determinant where all n electrons reside within the m orbitals is an initiator, all other determinants being noninitiators. During the simulation, however, the set of initiator *determinants* can also be dynamically enlarged so that if a noninitiator determinant acquires a population which exceeds a preset value (n_a), then it becomes an initiator determinant, i.e., acquires the ability to spawn progeny onto unoccupied determinants, and stays so as long as its population exceeds n_a. With this, the rigidity of the fixed CAS space is avoided and the initiator space is free to adopt any set of determinants, decided by the Hamiltonian of the system. This is a significant advantage when comparing to multireference (MR) methods.

Clearly, the above rules tend to the original, exact algorithm in three limits: first, as the number of walkers becomes very large, all determinants acquire a population and so all progeny survive, regardless of the state of the parent. Second, as we enlarge the fixed initiator space toward the full space, again all progeny will survive since all parent walkers are initiators. Finally, as n_a tends toward 0, all determinants are included in the initiator space. On the other hand, as we shall show below, the new rules effectively allow a dramatic reduction in the number of walkers.

We have tested this idea on the N_2 molecule in various basis sets. The N_2 molecule has been extensively studied as a prototype of a strongly correlated system, particularly as it is stretched to dissociation and has proven to be very challenging. N_2 is thus a stringent test for this method and accurate results indicate an ability to cope with many subsequent systems. This will be tested in studies yet to come in which a wide range of molecules are to be investigated. The FCI energy of N_2 is known in the cc-pVDZ basis, but the larger cc-pVTZ and cc-pVQZ basis sets are out of reach of conventional FCI calculations. However, these larger basis sets are needed for a proper description of the dynamical correlation.
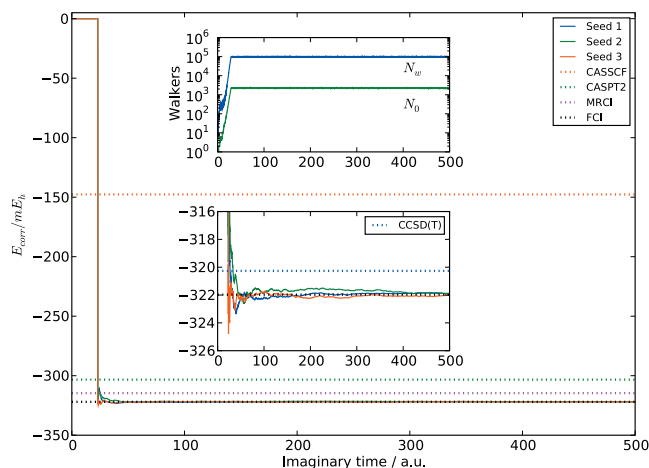

FIG. 1. The correlation energy ($E_{\text{corr}} = E - E_0$) from three simulations of N_2 at the equilibrium geometry, starting with different random number seeds, with a (10,8) initiator space. E was averaged after 23 a.u. of time had elapsed. Convergence rapidly occurs to values close to the shown FCI energy (Ref. 3): the lower inset is a fine-scale plot of E_{corr} in mE_h from the three simulations. Also shown are the CASSCF (Ref. 4), CASPT2 (Ref. 5), and MRCISD (Ref. 6) energies with the same (10,8) CAS space (Ref. 7) and CCSD(T) energy (Ref. 8). The upper inset shows N_w and N_0 over the course of the runs.

As a first example, we consider the N_2 molecule in the cc-pVDZ basis. In Ref. 1, we showed that the FCI energy of this system could be obtained with sub-mE_h accuracy but required 2.7×10^8 and 3.45×10^8 walkers at equilibrium and stretched geometries, respectively. By contrast, consider the evolution of the system with the additional rules starting with a (10,8) initiator space—which includes only 396 determinants—and setting $n_a = 3$. We started the simulation with one walker on D_0, and let the population grow until 100 000 walkers had been attained. After this, we changed to constant N_w mode with the shift varying and allowed the population dynamics to run for 5×10^5 time steps (of 0.001 a.u. imaginary time). During this run, the dynamic enlargement added a further $\langle N_{\text{ini}}\rangle = 2250$ determinants to the initiator space. Figure 1 shows the energy computed according to Eq. (1). Runs with different random number of seeds are shown, all of which converge to within sub-mE_h accuracy of the known exact result.

We also show the pattern of growth in both the total number of walkers and the number at D_0 ($\langle N_0 \rangle$). This behavior is markedly different from the original algorithm which in constant S mode would show an initial explosive (fast exponential) growth until the "annihilation plateau" is reached. This is followed by a period of time in this plateau, then a slower exponential rise whose rate is determined by the correlation energy. The population at D_0 typically would only begin to reach meaningful levels once in the second exponential growth phase. With the additional rules, however, the initial fast exponential growth phase and the annihilation plateau either disappears completely or is severely reduced. Instead we obtain essentially from the very beginning, a slow exponential growth in the number of walkers which resembles the second growth phase of the original algorithm. When the shift is allowed to vary, it quickly settles down to fluctuate around the correlation energy. The number of walkers at D_0 also grows exponentially from the

TABLE I. Summary of the simulation results for N_2 at equilibrium ($2.068a_0$) and stretched ($4.2a_0$) geometries in the VDZ basis set. $\langle N_{ini} \rangle$ is the equilibrium number of determinants added to the initiator spaces. The sizes of the fixed initiator spaces are 61, 396, 27×10^3, and 502×10^3 determinants, for the (10,7), (10,8), (10,11), and (10,14) CAS, respectively. $n_a = 3$ unless otherwise specified. The numbers in parentheses in the energies are the errors in the preceding digit, obtained from a blocking analysis (Ref. 9).

Initiator space	N_w	$N_c/10^3$	$\langle N_0 \rangle/10^3$	$\langle N_{ini} \rangle/10^3$	$(E - E_{FCI})/mE_h$
Equilibrium VDZ ($2.068a_0$)					
(10,7)	10^6	0	18	35	0.02(6)
(10,8)	10^4	0	0.330	0.22	2.2(8)
(10,8)	10^5	0	2.25	2.24	0.1(2)
$(10,8)_{n_a=3}$	10^6	0	17.2	33	0.01(7)
$(10,8)_{n_a=5}$	10^6	0	13.8	18.7	0.19(5)
$(10,8)_{n_a\to\infty}$	10^6	0	368	0	2.43(4)
(10,8)	10^7	0	154.0	349	0.02(3)
(10,11)	10^6	4.4	16.8	30.8	−0.07(9)
(10,14)	10^6	248	13.0	21	0.04(7)
CCSD					13.46
CCSD(T)					1.71
CASSCF (10,8)					174
CASPT2 (10,8)					18.7
MRCISD$_{(CASSCF\ ref.)}$ (10,8)					7.31
Stretched VDZ ($4.2a_0$)					
(10,7)	10^6	0	4	32	0.4(8)
(10,8)	10^4	0	0.08	0.14	16(5)
(10,8)	10^5	0	0.5	2.84	1(3)
(10,8)	10^6	0	4	29	−0.4(7)
(10,8)	10^7	0	40	403	0.05(60)
(10,11)	10^6	50	3.5	27	0.6(8)
(10,14)	10^6	630	1	12	1(1)
CASSCF (10,8)					185
CASPT2 (10,8)					12.4
MRCISD$_{(CASSCF\ ref.)}$ (10,8)					8.20

outset. This is important because the population at D_0 determines the accuracy to which the projected energy E of Eq. (1) can be measured. As a general rule, a simulation which achieves several thousand walkers on D_0 will obtain the correlation energy to within one part in 10^3 which is sub-mE_h accuracy for these systems. N_0 thus provides a measure of the total number of walkers required in the simulation.

We have looked at variations in both the fixed initiator size and the value of n_a, as well as different numbers of walkers, ranging from 10^4 to 10^7. The results are presented in Table I. The following can be seen. First, as the initiator spaces are enlarged, an annihilation plateau appears, and it is necessary to go to larger numbers of walkers before convergence can be achieved. However, the accuracy of the final result does not depend on the choice of fixed (CAS) initiator space. In fact, unnecessarily large spaces appear to be inefficient, as they make the accumulation of walkers at D_0 more difficult. Smaller fixed initiator spaces are therefore preferred and dynamic enlargement of the space, shown by $\langle N_{ini} \rangle$, becomes crucial to the accuracy. A very large n_a prevents any dynamic enlargement at all and determinants beyond single and double excitations of the initiator space may only be accessed via the special case of simultaneous spawning from two noninitiator determinants. The resulting energy is consequently too high. (Note if this special case is disallowed and $n_a \to \infty$, then in the limit of large N_w, this method becomes

comparable to MRCISD in a basis of HF orbitals). However, by reducing n_a until we no longer lower the energy, the point at which all significant determinants are being included in the initiator space may be deduced. The optimal value for this case appears to be $n_a = 3$, with which sub-mE_h accuracy is obtained with only a few thousand initiator determinants.

Also in Table I are presented results for stretched N_2, using the same parameters as above. It can be seen that this problem behaves similarly to the equilibrium geometry, except that generally somewhat more walkers are necessary to achieve convergence. Increasing numbers of walkers from 10^4 to 10^7, the error in the energy drops from $16mE_h$ to $0.05mE_h$, essentially by an order of magnitude for each order of magnitude increase in walker number. Two sources of error can be identified: a systematic error due to insufficient numbers of walkers and a stochastic error (as measured by blocking analysis) which can be reduced by running the simulations for longer. The results show that the systematic error is strongly reduced as the number of walkers is increased, and that one needs some 10^6 walkers to obtain sub-mE_h accuracy. Consistent with the discussion above, this corresponds to a few thousand walkers on D_0 and some 3×10^4 initiator determinants. It should be emphasized that although the computational demands of these calculations are one order of magnitude greater than at equilibrium geometry, this nevertheless represents a very cheap calculation

TABLE II. N_2 in cc-pVXZ ($X=3,4$) basis sets. Timesteps are $\tau=10^{-4}$ a.u. The fixed initiator space is (10,8) with $n_a=3$. The numbers in parentheses in the total energies are the errors in the preceding digit (Ref. 9).

		VTZ			VQZ		
	N_w	$N_0/10^3$	$\langle N_{ini}\rangle/10^3$	E_{tot}/E_h	$N_0/10^3$	$\langle N_{ini}\rangle/10^3$	E_{tot}/E_h
Equilibrium	10^5	0.85	1.3	$-109.373\ 9(2)$	0.6	0.85	$-109.400\ 1(2)$
($2.068a_0$)	10^6	5.5	14	$-109.375\ 2(2)$	3.3	9.761	$-109.404\ 8(2)$
	10^7	35	192	$-109.374\ 9(1)$	18.7	115	$-109.405\ 2(1)$
CCSD				$-109.355\ 35$			$-109.384\ 26$
CCSD(T)				$-109.373\ 70$			$-109.404\ 30$
CASSCF (10,8)				$-109.131\ 57$			$-109.139\ 44$
CASPT2 (10,8)				$-109.352\ 79$			$-109.384\ 52$
MRCISD$_{(CASSCF\ ref.)}$ (10,8)				$-109.361\ 09$			$-109.389\ 47$
Stretched	10^5	240	1.2	$-109.028(3)$	0.15	0.84	$-109.044(2)$
($4.2a_0$)	10^6	1.3	16	$-109.034(3)$	0.9	11	$-109.057\ 0(15)$
	10^7	8.9	188	$-109.042\ 1(5)$	4.7	130	$-109.061\ 2(8)$
CASSCF (10,8)				$-108.799\ 45$			$-108.799\ 45$
CASPT2 (10,8)				$-109.031\ 22$			$-109.031\ 22$
MRCISD$_{(CASSCF\ ref.)}$ (10,8)				$-109.025\ 36$			$-109.025\ 36$

(using about 25 Mbyte memory) to converge the FCI energy.

The experience developed above provides us with two useful simulation protocols. First, runs with different numbers of walkers should be done, monitoring the total energy for each. Convergence should be obtained in the sense that increasing walker number should not lead to a statistically significant lowering of the energy. Second, it is useful to monitor the number of walkers on D_0. In constant N_w mode, this number should ideally remain well above 10^3 while the projected energy is being averaged.

The results for the cc-pVTZ and cc-pVQZ basis sets are shown in Table II. The FCI spaces have 2.62×10^{12} and 1.55×10^{15} determinants, respectively (which is reduced by effectively a factor of two using time-reversal symmetry). For comparison we have also included CCSD, CCSD(T), CASSCF, CASPT2, and MRCISD results for equilibrium geometry, and CASSCF, CASPT2, and MRCISD at the stretched geometry (the coupled cluster methods diverge at this geometry). For both basis sets, we used the (10,8) fixed initiator spaces. At equilibrium geometry, simulations with 10^6 walkers result in an N_0 which exceed 10^3, and the corresponding energy can be calculated to sub-mE_h accuracy. Our predicted energies are between $1mE_h$ and $2mE_h$ below CCSD(T) which at equilibrium is a very good theory. At stretched geometry, we need some 10^7 walkers before N_0 substantially exceeds 10^3, but we still are able to obtain error bars in the sub-mE_h range. The CASSCF and CASPT2/MRCISD energies are higher in both cases.

Why does this simple algorithm prove so effective? We believe the reason is this: the new rule of "the survival of the fittest" means that whenever we add a new determinant to the system, it comes from a determinant whose sign has been well-established, either by virtue of being in a select group of CAS determinants, or by developing a sufficient population that makes it probable that its progeny will be sign-coherent as well. As a result the whole space can be sign-coherently sampled much more easily than with the original algorithm in which sign-coherence arose only when annihilation events became probable. This generally required a large number of walkers. It is remarkable that a very small number of determinants is able to endow sign-coherence to the whole space, even in the strongly multiconfigurational problem of stretched N_2. This is an extremely interesting feature of the algorithm, as once sign-coherence is achieved, the correct sampling of enormous FCI spaces becomes possible. It means that the dynamical correlation which large basis sets capture can be calculated with essentially FCI accuracy. This is a significant advance over MR perturbation methods such as CASPT2 which at present are the only viable way to treat such difficult problems. The algorithmic simplicity of the method is also an attractive feature. We will be examining the scaling properties of the new algorithm in a forthcoming paper, as well as applying it to other strongly correlated systems where both dynamic and static correlation are integral to the system.

[1] G. H. Booth, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **131**, 054106 (2009).
[2] It should be noted that mixed energy-estimators such as Eq. (1) are nonvariational.
[3] J. Olsen, O. Christiansen, H. Koch, and P. Jørgensen, J. Chem. Phys. **105**, 5082 (1996).
[4] B. O. Roos and P. R. Taylor, Chem. Phys. **48**, 157 (1980).
[5] K. Andersson, P.-A. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, J. Phys. Chem. **94**, 5483 (1990).
[6] H.-J. Werner and P. J. Knowles, J. Chem. Phys. **89**, 5803 (1988).
[7] H.-J. Werner, Mol. Phys. **89**, 645 (1996).
[8] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, and A. K. Chakraborty, Phys. Chem. Chem. Phys. **8**, 3172 (2006).
[9] H. Flyvbjerg and H. G. Petersen, J. Chem. Phys. **91**, 461 (1989).