12-7-2012

Semistochastic Projector Monte Carlo Method

F. R. Petruzielo

A. A. Holmes

Hitesh J. Changlani

M. P. Nightingale
University of Rhode Island, nightingale@uri.edu

C. J. Umrigar

Citation/Publisher Attribution

F. R. Petruzielo, A. A. Holmes, Hitesh J. Changlani, M. P. Nightingale, and C. J. Umrigar. (2012). "Semistochastic Projector Monte Carlo Method." *Physical Review Letters*, 109(23), 230201. doi: 10.1103/PhysRevLett.109.230201
Available at: http://dx.doi.org/10.1103/PhysRevLett.109.230201

Semistochastic Projector Monte Carlo Method

F. R. Petruzielo,[1,*] A. A. Holmes,[1,†] Hitesh J. Changlani,[1,‡] M. P. Nightingale,[2,§] and C. J. Umrigar[1,‖]

[1]*Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA*

[2]*Department of Physics, East Hall, University of Rhode Island, Kingston, Rhode Island 02881, USA*

(Received 1 May 2012; published 4 December 2012)

We introduce a *semistochastic* implementation of the power method to compute, for very large matrices, the dominant eigenvalue and expectation values involving the corresponding eigenvector. The method is semistochastic in that the matrix multiplication is partially implemented numerically exactly and partially stochastically with respect to expectation values only. Compared to a fully stochastic method, the semistochastic approach significantly reduces the computational time required to obtain the eigenvalue to a specified statistical uncertainty. This is demonstrated by the application of the semistochastic quantum Monte Carlo method to systems with a sign problem: the fermion Hubbard model and the carbon dimer.

DOI: 10.1103/PhysRevLett.109.230201 PACS numbers: 02.70.Ss, 31.10.+z, 31.15.V−, 71.15.−m

Introduction.—Consider the computation of the dominant eigenvalue of an $N \times N$ matrix, with N so large that the matrix cannot be stored. Transformation methods cannot be used in this case, but one can still proceed with the power method, also known as the projection method, as long as one can compute and store the result of multiplication of an arbitrary vector by the matrix. When, for sufficiently large N, this is no longer feasible, Monte Carlo methods can be used to represent stochastically both the vector and multiplication by the matrix. This suffices to implement the power method to compute the dominant eigenvalue and averages involving its corresponding eigenvector.

In this Letter, we propose a hybrid method consisting of numerically exact representation and multiplication in a small *deterministic* subspace, complemented by *stochastic* treatment of the rest of the space. This semistochastic projection method combines the advantages of both approaches: it greatly reduces the statistical uncertainty of averages relative to purely stochastic projection while allowing N to be large. These advantages are realized if one succeeds in choosing a deterministic subspace that carries a substantial fraction of the total spectral weight of the dominant eigenstate.

Semistochastic projection has numerous potential applications: transfer matrix [1] and quantum Monte Carlo (QMC) [2–4] calculations, respectively for classical statistical mechanical and quantum mechanical systems, and the calculation of subdominant eigenvalues [5].

In this Letter we apply the semistochastic method to compute the ground state energy of quantum mechanical Hamiltonians represented in a discrete basis. In this context, deterministic projection is known as full configuration interaction (FCI) to chemists and as exact diagonalization to physicists, whereas stochastic projection is the essence of various projector QMC methods [2,3]. Hence, semistochastic projection shall be referred to as the SQMC method. The benefit of the SQMC method over the corresponding QMC method is large in many systems of interest since the Hartree-Fock determinant, augmented by a small set of additional determinants, indeed represents a significant fraction of the total spectral weight of the ground state wave function.

The Hamiltonians for the systems considered here suffer from a sign problem; i.e., no sign changes of basis states can be found that render all off-diagonal matrix elements nonpositive (which allows all the coefficients of the desired eigenstate to be non-negative). Until recently, projector QMC had been used most successfully for systems that do not have a sign problem [2,3], or with an uncontrolled, variational fixed-node approximation [6]. The recent breakthroughs of Alavi and co-workers with their FCIQMC method [7] and its *initiator* extension [8], have enabled the treatment with a controllable bias of matrices with a sign problem. Consequently, the stochastic method to which we compare SQMC is essentially the same as the initiator FCIQMC (*i*-FCIQMC) method of Alavi with some minor differences as explained below.

Theory.—We start from an $N \times N$ Hermitian matrix H, with eigenvalues $E_0 < E_1 \leq \cdots \leq E_{N-1}$. In our case, H is a Hamiltonian represented in an orthonormal basis $\{|\phi_1\rangle, \ldots, |\phi_N\rangle\}$. To obtain the lowest eigenvalue E_0, and its eigenvector $\psi^{(0)}$ with components $\psi_i^{(0)} \equiv \langle\phi_i|\psi^{(0)}\rangle$, we first invert, shift, and scale the Hamiltonian matrix:

$$P = \mathbb{1} + \tau(E_T\mathbb{1} - H), \tag{1}$$

where E_T is a running estimate of E_0.

If $E_T = E_0$, P has unit eigenvalue. If $\tau < 2/(E_{N-1} - E_0)$, then the unit eigenvalue is the dominant one. With E_0 unknown, E_T is adjusted to ensure that the power method iterates remain reasonably constant in norm. When multiplication by P is performed deterministically, the fastest convergence rate is obtained for $\tau = 2/(E_{N-1} - E_1)$; semi- or fully-stochastic multiplications require smaller values of τ to reduce the statistical noise [9].

Let $\chi^{(0)}$ be an arbitrary initial vector satisfying $\langle\chi^{(0)}|\psi^{(0)}\rangle \neq 0$. Then, repeated application of P to $\chi^{(0)}$ yields

$$\chi^{(t+1)} = P\chi^{(t)} = P^{t+1}\chi^{(0)}. \tag{2}$$

According to the power method, $\chi^{(M)} \propto \psi^{(0)}$ for sufficiently large M. If coefficients $w_i^{(t)}$ are defined by the expansion

$$|\chi^{(t)}\rangle = \sum_{i=1}^{N} w_i^{(t)}|\phi_i\rangle, \tag{3}$$

the semistochastic representation of the weights $w_i^{(t)}$ and the multiplication by P in Eq. (2) are defined as follows.

Let $\mathcal{D}$ be the set of indices of vector components treated deterministically, and let $\mathcal{S}$ be the set of those treated stochastically, where $\mathcal{D} \cup \mathcal{S} = \{1, \ldots, N\}$, $\mathcal{D} \cap \mathcal{S} = \varnothing$, and $|\mathcal{D}| \ll N$. Accordingly, P is the sum of a deterministic block $P^{\mathcal{D}}$, and a stochastic complement $P^{\mathcal{S}}$,

$$P = P^{\mathcal{D}} + P^{\mathcal{S}}, \tag{4}$$

where

$$P_{ij}^{\mathcal{D}} = \begin{cases} P_{ij}, & \text{if } i, j \in \mathcal{D}, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

If the deterministic space is the entire space, then there is no sign problem or statistical noise. Consequently, we can expect that using a deterministic subspace that is not the entire space will reduce the sign problem and statistical noise.

The coefficients of the basis functions are represented as a population of walkers. The number of walkers on an occupied $|\phi_i\rangle$ is

$$n_i = \max(1, \lfloor |w_i| \rceil), \tag{6}$$

where $\lfloor\cdot\rceil$ denotes the nearest integer and each walker has signed weight w_i/n_i.

Next, we proceed to the multiplication by P which evolves the coefficients from time t to time $t+1$.

(i) To account for the off-diagonal elements in $P^{\mathcal{S}}$, for each walker on $|\phi_i\rangle$, a move to $|\phi_j\rangle \neq |\phi_i\rangle$ is made with probability T_{ji}. A single walker on $|\phi_i\rangle$ contributes

$$\begin{cases} 0, & \text{if } i, j \in \mathcal{D}, \\ \dfrac{P_{ji}}{T_{ji}} \dfrac{w_i^{(t)}}{n_i^{(t)}}, & \text{otherwise} \end{cases} \tag{7}$$

to the signed walker weight on $|\phi_j\rangle$. The choice of T determines the probability that particular off-diagonal moves are made. In this Letter, the near-uniform choice of Booth, Thom, and Alavi is used [7]. To control sign problems present in our examples, we use the initiator idea [8], which we generalized in that we increase the initiator threshold with the number of steps taken since the last visit to the deterministic space [10].

(ii) To account for the diagonal elements in $P^{\mathcal{S}}$, the contribution to the total signed walker weight on $|\phi_j\rangle$, with $j \in \mathcal{S}$, is

$$P_{jj} w_j^{(t)}. \tag{8}$$

(iii) Deterministic evolution is performed with $P^{\mathcal{D}}$. The contribution to the signed weight on $|\phi_j\rangle$, with $j \in \mathcal{D}$, is

$$\sum_{i \in \mathcal{D}} P_{ji}^{\mathcal{D}} w_i^{(t)}. \tag{9}$$

$P^{\mathcal{D}}$ is stored and applied as a sparse matrix.

(iv) Finally, for each $|\phi_j\rangle$, all signed walker weight generated on $|\phi_j\rangle$ is summed, taking into account the sign of the contribution. To avoid the large computational and memory cost of having small weights on a large number of basis states, basis states with weight less than some minimum cutoff, $w_{\min}$, are combined via an unbiased prescription [11].

After sufficiently many multiplications by P, contributions from subdominant eigenvectors die out on average. At this point, the collection of averages begins. The most commonly employed estimator for the dominant eigenvalue is the *mixed estimator*

$$E_{\text{mix}} = \frac{\langle\psi^{(0)}|\hat{H}|\psi_T\rangle}{\langle\psi^{(0)}|\psi_T\rangle}, \tag{10}$$

where the trial state $|\psi_T\rangle$ satisfies $\langle\psi^{(0)}|\psi_T\rangle \neq 0$.

The trial state $|\psi_T\rangle$ is a linear combination of basis states [12],

$$|\psi_T\rangle = \sum_{i \in \mathcal{T}} d_i|\phi_i\rangle, \tag{11}$$

where $\mathcal{T}$ is the set of indices of those basis functions that contribute to the trial state. We require that $|\mathcal{T}| \ll N$, but not necessarily that $\mathcal{T} \subset \mathcal{D}$.

At any particular time t, the stochastic representation of the dominant eigenvector is

$$|\psi^{(0)}\rangle \approx |\chi^{(t)}\rangle = \sum_{i \in \mathcal{W}^{(t)}} w_i^{(t)}|\phi_i\rangle, \tag{12}$$

where $\mathcal{W}^{(t)}$ is the set of indices of basis functions occupied by walkers at time t. The full representation of the dominant eigenvector is obtained by averaging over Monte Carlo generations

$$|\psi^{(0)}\rangle \approx \frac{1}{N_{\text{gen}}} \sum_{t=1}^{N_{\text{gen}}} \sum_{i \in \mathcal{W}^{(t)}} w_i^{(t)}|\phi_i\rangle, \tag{13}$$

where N_{gen} is the number of times P is applied after equilibration.

For the trial state in Eq. (11), E_{mix} of Eq. (10) is

$$E_{\text{mix}} = \frac{\sum_{t=1}^{N_{\text{gen}}} \sum_{i \in \mathcal{W}^{(t)}} w_i^{(t)} \sum_{j \in \mathcal{T}} H_{ij} d_j}{\sum_{t=1}^{N_{\text{gen}}} \sum_{i \in \mathcal{W}^{(t)} \cap \mathcal{T}} w_i^{(t)} d_i}. \quad (14)$$

Since E_{mix} is a zero-variance and zero-bias estimator when $|\psi_T\rangle$ is equal to the dominant eigenvector, improving the quality of $|\psi_T\rangle$ reduces fluctuations and bias in the mixed estimate of the dominant eigenvalue. This reduction can be achieved with almost no additional computational cost by storing nonzero $\sum_{j \in \mathcal{T}} H_{ij} d_j$ terms.

The trial wave function space and the deterministic space are generated with identical iterative schemes, but possibly different parameters. At each iteration, first define a reference space as all states obtained in the previous iteration. Second, generate a space which includes all determinants connected to the reference space by a single application of the Hamiltonian. Third, find the dominant eigenvector in this space. Fourth, truncate the space using a criterion based on the magnitude of the coefficient of each state in the eigenvector. This truncated space becomes the reference for the next iteration. The reference for the first iteration is the Hartree-Fock state.

Applications.—The semistochastic method is now applied to compute the ground state energy of the carbon dimer and the simple-square 8×8 fermionic Hubbard model with periodic boundaries. In both cases, we represent H in the basis of determinants formed from the restricted Hartree-Fock orbitals. For the Hubbard model these orbitals are the momentum eigenstates. For the carbon dimer these orbitals are obtained by solving the restricted Hartree-Fock equations in cc-pVTZ basis set [13]. The majority of the Hubbard calculations are performed for $U/t = 4$, where U is the on-site Coulomb repulsion and t is the nearest neighbor hopping parameter. This parametrization is considered to be in the *intermediate coupling* regime (the noninteracting bandwidth being $8t$), and has been used widely in the literature [14].

For various sizes of the deterministic space, we demonstrate the improvements of the SQMC method over the purely stochastic method defined by a deterministic space which includes only the Hartree-Fock determinant. The purely stochastic method is almost the same as the *i*-FCIQMC method [7,8], aside from some details such as the use of real walker weights versus the integer walker weights used in the FCIQMC method and the use of a graduated initiator in the SQMC method [10]. The most dramatic benefit of the SQMC method is in the efficiency, which is defined to be proportional to the inverse of the time required to obtain the ground state energy to a specified level of uncertainty.

To show the gain in efficiency of the SQMC method we computed the relative efficiency, i.e., the efficiency

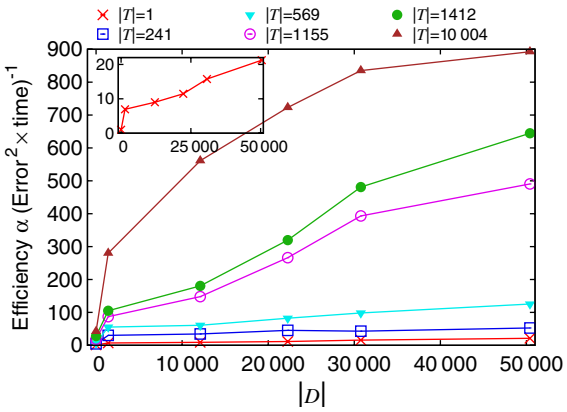

FIG. 1 (color online). Relative efficiency of the SQMC method vs dimension $|\mathcal{D}|$ of the deterministic space for the simple-square 8×8 Hubbard model with periodic boundaries, $U/t = 4$ and 10 electrons. Results are shown for trial wave functions of increasing size. The inset shows the $|\mathcal{T}| = 1$ curve on an expanded scale. For this system, $N \approx 10^{12}$.

normalized by that of the stochastic method ($|\mathcal{D}| = 1$), with $|\mathcal{T}| = 1$. Figure 1 shows the relative efficiency of the SQMC method vs the size of the deterministic space for the simple-square 8×8 Hubbard model with periodic boundaries, $U/t = 4$ and 10 electrons. The orders of magnitude increases in efficiency demonstrate the benefits not only of the SQMC method but also of improving the trial wave function. The gain of just using the largest deterministic space is a factor of 22, while the benefit of just using the largest trial wave function is a factor of 42. Both together yield a factor of about 900 as seen in the plot, but the two are not always multiplicative.

Figure 2 shows the efficiency gain of the SQMC method vs filling fraction for the simple-square 8×8 Hubbard model with $U/t = 4$. The deterministic space, constructed by applying the Hamiltonian once to the Hartree-Fock determinant, has a rather modest increase in size from 1412 to 16 540 determinants, whereas the size of the Hilbert space grows enormously from about 10^{12} to 10^{35}. Nevertheless, the efficiency gains increase with filling fraction. Calculations beyond the scope of the present paper show that the initiator bias, at all fillings, decreases with increasing $\mathcal{D}$, but that it increases with filling fraction and U in both the stochastic and the semistochastic methods.

The SQMC method produces large efficiency gains for chemical systems as well. Figure 3 shows the efficiency gain of SQMC vs the size of the deterministic space for the carbon dimer with a cc-pVTZ basis set [13]. The bottom two curves are for $\mathcal{D}$ and $\mathcal{T}$ generated with one applications of our iterative scheme which generate single and double excitations only. The largest efficiency gain for these is about 40. The top two curves are for $\mathcal{D}$ and $\mathcal{T}$ generated with two applications of our iterative scheme and, hence, include several chemically relevant quadruple excitations which are important for correctly describing

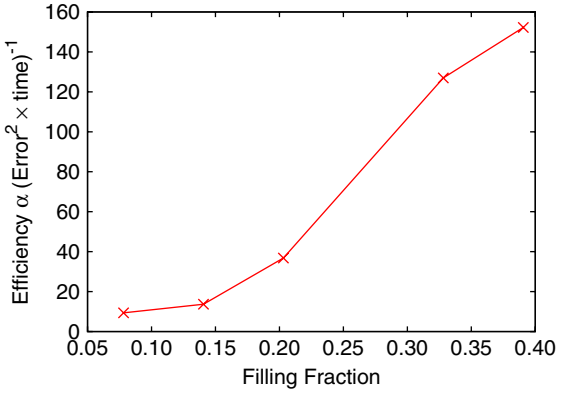

FIG. 2 (color online). Relative efficiency of the SQMC method vs filling fraction for the simple-square 8×8 Hubbard model with $U/t = 4$. In all cases, the trial wave function is the Hartree-Fock determinant. The deterministic space is constructed by applying the Hamiltonian once to the Hartree-Fock determinant. This yields spaces of sizes 1412, 4088, 7424, 14 160, 16 540. N ranges from roughly 10^{12} to 10^{35}. Nevertheless, the efficiency gains increase with filling fraction. Calculations beyond the scope of the present paper show that the initiator bias, at all fillings, decreases with increasing $\mathcal{D}$, but that it increases with filling fraction and U in both the stochastic and the semistochastic methods.

the ground state wave function. The largest efficiency gain now jumps to over 1000.

Not only is the SQMC method much more efficient than the stochastic method, but in some cases, also the initiator bias is significantly reduced. Figure 4 shows the biased estimates of the energy as obtained by the semistochastic and stochastic methods vs the average number of occupied determinants for the 8×8 Hubbard model with $U/t = 1$

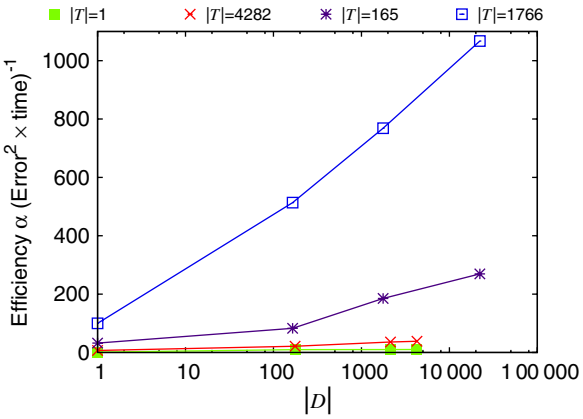

FIG. 3 (color online). Relative efficiency of SQMC vs dimension $|\mathcal{D}|$ of the deterministic space for the carbon dimer with a cc-pVTZ basis. Results are shown for trial wave functions of increasing size. The top two curves are for $\mathcal{D}$ and $\mathcal{T}$ generated with two applications of our iterative scheme. The 165 and 1766 determinant wave functions with some quadruple excitations have much higher efficiency than the 4282 determinant wave function without any. For this system, $N \approx 10^9$.

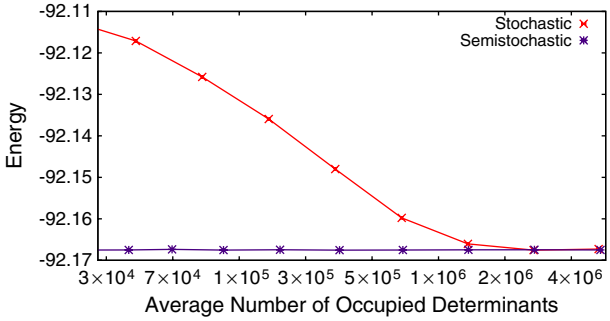

FIG. 4 (color online). Energy obtained by the semistochastic and stochastic methods vs the average number of occupied determinants for the simple-square 8×8 Hubbard model with $U/t = 1$ and 50 electrons. The trial wave function for each of these calculations is the Hartree-Fock determinant. The deterministic space consists of the 16 540 determinants connected to the Hartree-Fock determinant. For this system, $N \approx 10^{35}$.

and 50 electrons. The SQMC method has essentially no bias. A larger average number of occupied determinants corresponds to using a larger walker population in the calculation. The time required for a step in the calculation is proportional to the walker population.

The reduction in initiator bias is not always large. Figure 5 shows both the SQMC and stochastic method energy vs the average number of occupied determinants for the 8×8 Hubbard model with $U/t = 4$ and 10 electrons. The SQMC method has a reduced initiator bias for a small, but not for a large number of occupied determinants. However, for this system and all other systems studied, the SQMC method has a smoother bias than the stochastic method.

Conclusion.—The semistochastic power method, a hybrid with deterministic and stochastic components, was introduced for finding the dominant eigenvalue and sampling the corresponding eigenvector of a matrix. We showed that this novel, deterministic component

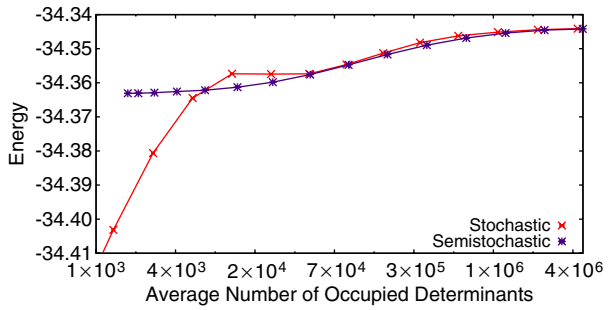

FIG. 5 (color online). Energy obtained by the semistochastic and stochastic methods vs the average number of occupied determinants for the simple-square 8×8 Hubbard model with $U/t = 4$ and 10 electrons. The trial wave function for each of these calculations is the Hartree-Fock determinant. The deterministic space for the SQMC method consists of the 1412 determinants connected by the Hamiltonian to the Hartree-Fock determinant. For this system, $N \approx 10^{12}$.

significantly reduces the noise of the purely stochastic method without compromising its ability to deal with matrices well beyond the size that can be handled by purely deterministic methods. In particular, matrices ranging in order from 10^9 to 10^{35} were successfully tackled. Besides being more efficient than a purely stochastic approach, the semistochastic method has in some cases the additional benefit of a much reduced initiator bias. Also, the bias tends to be smoother and more amenable to removal by extrapolation. We only presented applications to systems with a sign problem, but the efficiency benefits of a semistochastic implementation of the power method extend to systems without a sign problem.

We thank Garnet Chan and Ali Alavi for valuable discussions. This work was supported in part by DOE-CMCSN DE-SC0006650 (AAH), NSF CHE-1112097 (HJC) and NSF DMR-0908653 (FRP).

*frp3@cornell.edu
†aah95@cornell.edu
‡hjc55@cornell.edu
§nigh@pobox.com
‖CyrusUmrigar@cornell.edu

[1] M. P. Nightingale and H. W. J. Blöte, Phys. Rev. Lett. **60**, 1562 (1988).
[2] R. Blankenbecler and R. L. Sugar, Phys. Rev. D **27**, 1304 (1983).
[3] N. Trivedi and D. M. Ceperley, Phys. Rev. B **40**, 2737 (1989); **41**, 4552 (1990).
[4] *Quantum Monte Carlo Methods in Physics and Chemistry*, NATO ASI Ser. C Vol. 525, edited by M. P. Nightingale and C. J. Umrigar (Kluwer, Dordrecht, 1999).
[5] M. P. Nightingale and H. W. J. Blöte, Phys. Rev. B **62**, 1089 (2000).
[6] H. J. M. van Bemmel, D. F. B. ten Haaf, W. van Saarloos, J. M. J. van Leeuwen, and G. An, Phys. Rev. Lett. **72**, 2442 (1994); D. F. B. ten Haaf, H. J. M. van Bemmel, J. M. J. van Leeuwen, W. van Saarloos, and D. M. Ceperley, Phys. Rev. B **51**, 13039 (1995).
[7] G. H. Booth, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **131**, 054 106 (2009).
[8] D. Cleland, G. H. Booth, and A. Alavi, J. Chem. Phys. **132**, 041 103 (2010).
[9] For each system studied, a near-optimal value of τ can be determined from inexpensive calculations with a small number of walkers.
[10] In the initiator approach, a state must have a minimum absolute weight i in order to contribute an off-diagonal move to a state that is not already occupied. This results in a nonvariational bias that vanishes in the limit of infinite walker population. In our modification of this approach, $i = cm^p$, where m is the number of moves the walker has made since its last visit to the deterministic space and c and p are constants, chosen to be 1 for the applications in this Letter.
[11] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, J. Chem. Phys. **99**, 2865 (1993).
[12] In FCIQMC [7,8], a single state, the Hartree-Fock determinant, has been used as the trial state.
[13] T. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[14] For example, J. E. Hirsch, Phys. Rev. B **31**, 4403 (1985); N. Furukawa and M. Imada, J. Phys. Soc. Jpn. **61**, 3331 (1992); H. D. Raedt and M. Frick, Phys. Rep. **231**, 107 (1993); S. Zhang, J. Carlson, and J. E. Gubernatis, Phys. Rev. B **55**, 7464 (1997).