Making the most of data: Quantum Monte Carlo postanalysis revisited

Tom Ichibha[*]
School of Information Science, JAIST, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan

Verena A. Neufeld
Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom and Department of Chemistry, Columbia University, New York, New York 10027, USA

Kenta Hongo
Research Center for Advanced Computing Infrastructure, JAIST, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan

Ryo Maezono[†]
School of Information Science, JAIST, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan

Alex J. W. Thom[‡]
Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

(Received 1 June 2021; accepted 10 March 2022; published 19 April 2022)

In quantum Monte Carlo (QMC) methods, energy estimators are calculated as (functions of) statistical averages of quantities sampled during a calculation. Associated statistical errors of these averages are often estimated. This error estimation is not straightforward and there are several choices of the error estimation methods. We evaluate the performance of three methods (the Straatsma method, an autoregressive model, and a blocking analysis based on von Neumann's ratio test for randomness) for the energy time series given by three QMC methods [diffusion Monte Carlo, full configuration interaction Quantum Monte Carlo (FCIQMC), and coupled cluster Monte Carlo (CCMC)]. From these analyses, we describe a hybrid analysis method which provides reliable error estimates for a series of various lengths of FCIQMC and CCMC's time series. Equally important is the estimation of the appropriate start point of the equilibrated phase. We establish that a simple mean squared error rule method as described by White [K. P. White, Jr., Simulation **69**(6), 323 (1997)] can provide reasonable estimations.

DOI: 10.1103/PhysRevE.105.045313

I. INTRODUCTION

With the increase in availability of large-scale computers, quantum Monte Carlo (QMC) methods are more feasible now owing to the good parallelizability of such algorithms [1–3]. QMC methods are generally among the most accurate *ab initio* methods and are often used for systems which cannot be described by density functional theory with sufficiently reliable accuracy [4–7] or which are too large for conventional post Hartree-Fock methods (coupled cluster, full-configuration interaction, etc.) [8].

Diffusion Monte Carlo (DMC) [9] is such a QMC method with a computational scaling of $O(N^3)$ for a system of N electrons and, as such, it can be applied to large-size systems including more than 1000 electrons [5]. The drawback of DMC is the requirement to use the fixed-node or fixed-phase approximation [10–12] to avoid the sign problem, which introduces a bias dependent on the quality of the nodes of a trial wave function. The methods to alleviate this error [13,14] make calculations considerably more expensive.

Some newer QMC methods in quantum chemistry which are not constrained by the fixed-node approximation have attracted interest of late. One is the full-configuration interaction QMC (FCIQMC) method [15–18], which stochastically solves the equations of full-configuration interaction (FCI) by sampling with discrete particles. Although the scaling of the calculation cost is still exponential [18] in the number of electrons like FCI, the prefactor of the scaling curve is significantly reduced. Thus this method is applicable to moderate-size molecules, e.g., tetramethyleneethane (C_6H_8) [19], while FCI has been used for small molecules just with two or three atoms in most cases. Another related method is coupled cluster Monte Carlo (CCMC), which stochastically solves coupled cluster (CC) equations [20,21]. Since the parameter space of a truncated CC calculation is smaller than that of FCI, CCMC generally has a smaller memory cost than FCIQMC.

QMC methods commonly estimate energies as (functions of) statistical averages of quantities sampled during a calculation, along with an associated statistical error. It is nontrivial to estimate the error reliably due to the following factors

[*]ichibha@icloud.com
[†]rmaezono@mac.com
[‡]ajwt3@cam.ac.uk

[22]: (i) autocorrelation between samples and (ii) heavy tails of the sampling distribution. In this work, we examine the performance of three characteristic automatic error estimation methods: Straatsma [23], the autoregressive (AR) model [24], and blocking analysis with block size chosen based on von Neumann's ratio test for randomness [25,26] (we call it von Neumann blocking) for the energy time series obtained by applying DMC, FCIQMC, and CCMC to simple systems. DMC was applied to a (1) Ne atom, (2) N_2 molecule, and (3) spin nonpolarized electron gas. FCIQMC was applied to (1) and (3). CCMC was applied to (1) and (2). From these analyses, we establish recommendations for the most reliable error estimation method for different lengths of data or different normality of distribution. We find that the difference in performance of the error estimation methods is pronounced when the length of the data is small, and the methods tend to underestimate the error. We devise a hybrid scheme working well for any length of time series generated by FCIQMC and CCMC.

Another important issue on the postanalysis of QMC is to determine the length of the equilibration (warm-up) phase [22]. Underestimation of the length gives a bias in the energy, but its overestimation also increases the statistical error. In this work, we evaluated a simple heuristic method called the mean squared error rule (MSER) [27] and established that the method provided reasonable estimates in our tests.

II. ERROR ESTIMATION METHODS

In this section, we explain the error estimation methods tested: The Straatsma method [23], the AR model [24], and the von Neumann blocking [26]. In this section, we assume that the central limit theorem is satisfied: The probability distribution of the mean is the normal distribution.

A. Straatsma

Straatsma *et al.* calculate the variance of the statistical mean of the stationary time series by

$$\sigma_{\bar{X}}^2 = s_X^2 \cdot \tau / n \tag{1}$$

[23]. Here, τ is the autocorrelation length among the steps (given later), n is the number of steps in the time series, and s_X^2 is the standard variance of $\{X_i\}$ under an assumption that the samples are independent of each other:

$$s_X^2 \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{2}$$

The autocorrelation length τ is given as a sum of autocorrelation functions $\{c_\lambda\}$,

$$\tau \equiv 1 + 2\sum_{\lambda=1}^{n-1}\left(1 - \frac{\lambda}{n}\right)c_\lambda. \tag{3}$$

The autocorrelation function with lag λ, c_λ, is approximately given with the finite number n of samples as

$$c_\lambda \approx \frac{1}{s_X^2 \cdot (n-\lambda)}\sum_{i=1}^{n-\lambda}\{(X_i - \bar{X})(X_{i+\lambda} - \bar{X})\}. \tag{4}$$

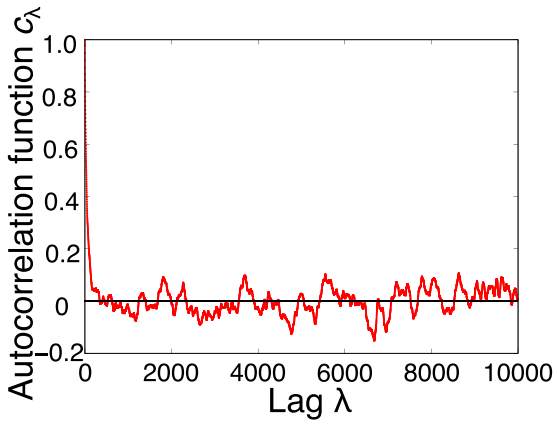

FIG. 1. A typical autocorrelation function of the energy time series generated by the CCMC method applied to the Ne atom. It rapidly decreases and oscillates around zero as the lag λ increases, showing that the correlation between X_i and $X_{i+\lambda}$ decreases.

This approximation becomes inaccurate for large λ since the number of terms to be summed up is small. Thus, in Eq. (3), we limit the summation to values before c_λ becomes negative for the first time, since later c_λ oscillates around zero (see Fig. 1). The resulting τ is then used to calculate $\sigma_{\bar{X}}^2$ in Eq. (1). The validity of the criterion is discussed in Appendix A.

B. Autoregressive (AR) model

The AR model assumes that the given random sampling $\{X_i\}_{i=1}^n$ can be reasonably described by [24]

$$X_i = \bar{X} + \eta_1 X_{i-1} + \eta_2 X_{i-2} + \cdots + \eta_p X_{i-p} + a_i, \tag{5}$$

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n). \tag{6}$$

The ith sample is given as a linear combination of the previous steps and a random Gaussian noise a_i with average 0 and variance σ_a^2. The coefficients $\{\eta_i\}_{i=1}^p$ and the variance σ_a^2 are fitted to the given time series with a least-squares procedure, which leads to the so-called Yule-Walker equations [28]. The number of coefficients p in the AR model is decided based on Akaike's information criterion (AIC) [29]. AIC decides the maximum likelihood p value by minimizing

$$\mathrm{AIC} = n\left[\ln\left(2\pi\frac{\mathrm{SSE}}{n}\right) + 1\right] + 2(p+1). \tag{7}$$

Here, SSE is the sum of the squared deviations between the actual X_i and one estimated by the model. Hence the first term of AIC evaluates how the model fits the given time series well, whereas the latter term plays a role to avoid an overfitting with an excessive number of parameters. The estimation of the autocorrelation length τ is calculated by

$$\tau = \left(1 - \sum_{\lambda=1}^{p} c_\lambda \eta_\lambda\right) \Bigg/ \left(1 - \sum_{\lambda=1}^{p}\eta_\lambda\right)^2, \tag{8}$$

where c_λ are defined by Eq. (4).

C. von Neumann blocking

At first, von Neumann blocking applies a blocking operation to the original time series and judges if the distribution of the means of the values in the blocks is regarded to be an independent and identical normal distribution (i.i.d.) using the von Neumann's ratio test for randomness [25,26]. The block length is increased until the test is passed. This process is to guarantee a normal distribution for the mean of the blocks.

We explain this method with equations. First, the original time series $\{X_i\}_{i=1}^{n}$ is divided into blocks,

$$W_j(m) = \frac{1}{m}\sum_{l=1}^{m} X_{m(j-1)+l} \ (j = 1, 2, 3, \dots), \tag{9}$$

where m is the length of each block. To apply the ratio test for randomness, we need to compute two quantities: The variance of $\{W_j(m)\}$, assuming samples are independent of each other,

$$s_W^2(m) = \frac{1}{k-1}\sum_{j=1}^{k}[W_j(m) - \bar{W}(m)]^2, \tag{10}$$

$$\bar{W}(m) = \frac{1}{k}\sum_{j=1}^{k} W_j(m), \tag{11}$$

and a mean square successive difference,

$$\mathrm{MSSD}_W(m) = \frac{1}{k-1}\sum_{j=1}^{k-1}[W_{j+1}(m) - W_j(m)]^2, \tag{12}$$

where k is the number of blocks, defined as $k = \mathrm{floor}(n/m)$. Then, evaluate the von Neumann test statistic,

$$c_m = 1 - \frac{\mathrm{MSSD}_W(m)}{2s_W^2(m)}. \tag{13}$$

When the data points in the time series $\{W_j\}_{i=1}^{k}$ are i.i.d., the distribution of statistic c_m asymptotically approaches the standard normal distribution with $k \to \infty$. Therefore, the null hypothesis, i.e., "the times series $\{W_j\}_{i=1}^{k}$ is regarded to follow i.i.d.," is accepted with a risk factor α when the statistic c_m is smaller than the $\alpha/2$ quantile of the left-sided standard normal distribution. A disadvantage of this method is that the test for randomness is not accurate when the number of blocks k is small. Therefore, we restrict the number of blocks to be at least 10 even if the test is not passed, following previous work [26].

III. MSER METHOD

MSER [27,30] aims to give an adequate estimate of the warm-up steps so as to reduce the bias from the warm-up phase and the statistical error in a proper balance. The number of warm-up steps d is determined by minimizing the standard error in the mean after a step d,

$$\mathrm{MSER}(d) = \frac{s_X^2(d)}{n-d}, \tag{14}$$

$$s_X^2(d) \equiv \frac{1}{n-d}\sum_{i=1}^{n-d}(X_{i+d} - \bar{X})^2, \tag{15}$$

$$\bar{X} = \frac{1}{n-d}\sum_{i=1}^{n-d} X_{i+d}. \tag{16}$$

Here, $s_X^2(d)$ is the population variance of X_i, which oscillates around a constant value with d if the time series is already equilibrated at the dth step. When a warm-up phase is present, $s_X^2(d)$ increases from the constant value according to how much the warm-up phase remains in $\{X_i\}_{i=d+1}^{n}$. Meanwhile, $1/(n-d)$ monotonically increases according to d. The MSER method decides the number of warm-up steps as the d minimizing their product, i.e., MSER(d) in Eq. (14). We have increased the d from 0 to $0.9n$ in $0.01n$ increments and surveyed what d minimizes the MSER(d).

IV. QMC CALCULATION DETAILS

We employed CASINO [31] for the DMC calculations. We used the trial wave functions provided in the examples directory of CASINO for the Ne atom and the N_2 molecule. They are Slater-Jastrow-type wave functions [9]. The determinant part is generated by the Hartree-Fock method implemented in CRYSTAL03 [32]. The orbitals are expanded by the Gaussian basis set. The numerical details of the basis sets are found in the CRYSTAL03 input files attached as the Supplemental Material [33]. The Jastrow factor consists of one- and two-body terms and includes 42 parameters for the Ne atom and 72 parameters for the N_2 molecule. For the DMC calculations, the target population of walkers is set to be 1024 for both and the time step is 0.005 a.u.$^{-1}$ for the Ne atom and 0.01 a.u.$^{-1}$ for the N_2 molecule, respectively. We used the smoothly cut-off drift vector [34] to reduce the time-step error. We also used the cusp correction scheme [35], which applies an additive cusp-enforcing function to the relevant part of Gaussian orbitals to satisfy the Kato cusp conditions [36]. The electron gas determinental wave function was constructed from plane waves. The density of the electron gas is set to one electron per sphere with a radius of 1 Bohr, i.e., r_s=1 Bohr, and the simulation cell includes 54 electrons in total. The Jastrow factor includes only two-body terms, with a total of eight parameters. Each sample of the energy time series is given by averaging the local energies [9] over all of the walkers for every QMC iteration.

We performed FCIQMC and CCMC calculations with HANDE [37,38]. For the Ne atom and the N_2 molecule, the reference Slater determinant is prepared by the Hartree-Fock method with the cc-pVDZ Gaussian basis set [39] using PSI4 [40]. The target number of the walker population for calculations other than the CCMC/N_2 molecule is 500 and that for the CCMC/N_2 molecule is 4000 walkers (walkers in FCIQMC and CCMC are also called psips and excips, respectively). The walker population for CCMC was decided based on the shoulder plot to be large enough to achieve a stable population [21]. The time step is 2.0×10^{-5} a.u.$^{-1}$, and we took 100 decorrelation steps between samples of the energy. Next, in the case of the electron gas, the electronic density corresponds to $r_s = 1$ Bohr and the simulation cell includes six electrons in total. The target population is 2000 and the time step is 0.02 a.u.$^{-1}$ for FCIQMC and 0.01 a.u.$^{-1}$ for CCMC, respectively. The excitation operator of CCMC consists of single- and double-excitation operators, so the predicted energy corresponds to the coupled cluster with single and double excitations (CCSD) method.

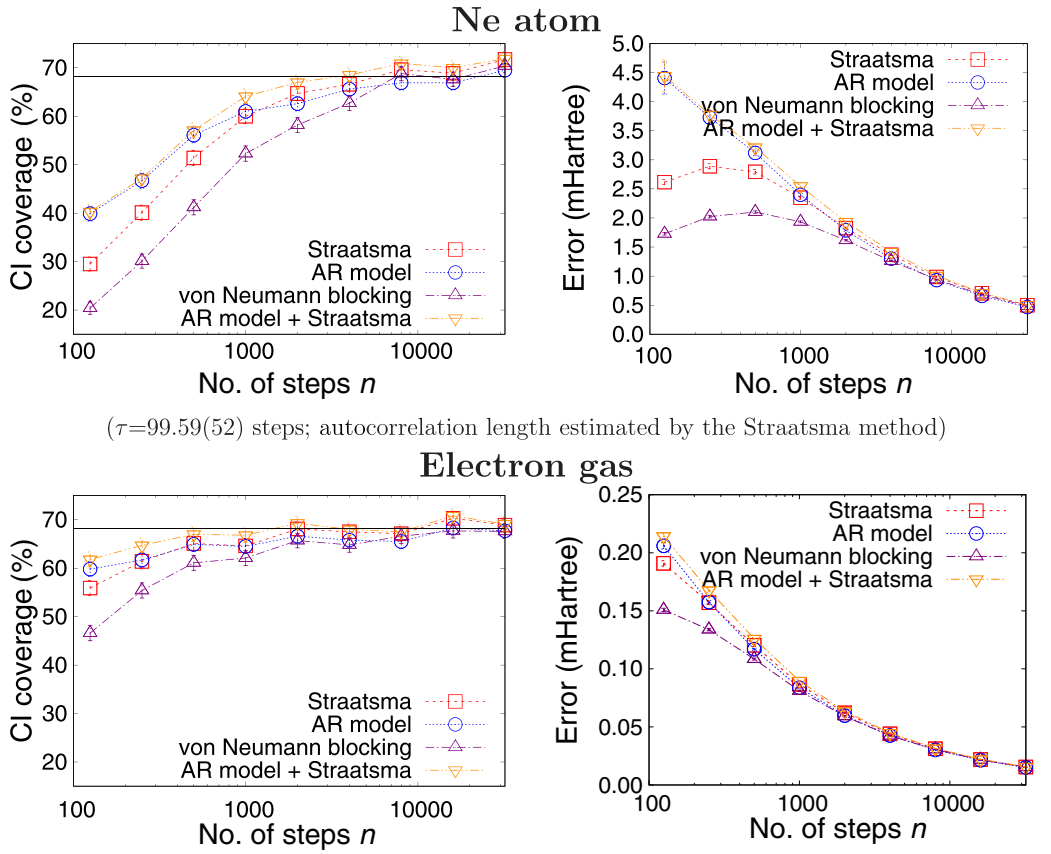

FIG. 2. Plots of CI coverage and statistical error for the time series given by FCIQMC. The left and right columns show the estimated CI coverage and the statistical error of the energy mean given by applying the error estimation methods to the time series of the Ne atom and the electron gas. The autocorrelation lengths τ are calculated by Straatsma.

Each sample of the energy time series is given as the instantaneous projected energy,

$$E = \langle D_0|\hat{H}|\Psi\rangle / \langle D_0|\Psi\rangle, \tag{17}$$

for every QMC iteration. Here, $\hat{H}$ is a Hamiltonian, D_0 is a determinant of the reference state (usually the Hartree-Fock determinant), and Ψ is an instantaneous wave function composed of multiple determinants,

$$|\Psi\rangle = c_0|D_0\rangle + \sum_{a,i} c_i^a \left|D_i^a\right\rangle + \sum_{a,b,i,j} c_{i,j}^{a,b} \left|D_{i,j}^{a,b}\right\rangle + \cdots. \tag{18}$$

Here, D_i^a and $D_{i,j}^{a,b}$ are determinants, where a (and b) instead of i (and j) are now occupied, respectively. With the simulation time evolution, the coefficients c_i^a, $c_{i,j}^{a,b}$, ... converge to a solution proportional to the FCI (CC) solution in FCIQMC (CCMC). Substituting $|\Psi\rangle$ of Eq. (18) into Eq. (17), the instantaneous projected energy is given as

$$E = \langle D_0|\hat{H}|D_0\rangle + \sum_{a,i} \frac{c_i^a}{c_0} \langle D_0|\hat{H}\left|D_i^a\right\rangle + \sum_{a,b,i,j} \frac{c_{i,j}^{a,b}}{c_0} \langle D_0|\hat{H}\left|D_{i,j}^{a,b}\right\rangle + \cdots. \tag{19}$$

The energy mean and error estimation using the instantaneous projected energy explained above is different from the commonly used analysis using the projected energy estimator [15,20]. In this analysis, the cross-correlation function of $\langle D_0|\hat{H}|\Psi\rangle$ and $N_0 \equiv \langle D_0|\Psi\rangle$ are used for the error estimation, following Wolff's theory [41]: This treatment works to eliminate the bias stemming from the correlation between the two variables in theory. However, we did not observe such a bias when using the instantaneous projected energy for the error estimation in the case of a Ne atom. We estimated the correlation energy and its error by the hybrid method applied to 1000 time series of the instantaneous projected energy. Each time series was generated by the CCMC method and the length was 200 000, where the equilibration steps were already excluded. We compared the correlation energies of 1000 time series to the deterministic CCSD value, −0.190 861 Hartree, calculated with PSI4 [40]. As a result, the mean of the absolute differences is $3.0(1)\times10^{-5}$ Hartree and the maximum absolute difference is $14(19)\times10^{-5}$ Hartree that is identical to zero within the error bar: The bias caused by using the instantaneous projected energy is not expected to be significant in this analysis.

V. RESULTS AND DISCUSSION

A. Confidence interval coverage

We explain how we evaluate the performance of the error estimation methods. First we prepared 1000 different energy

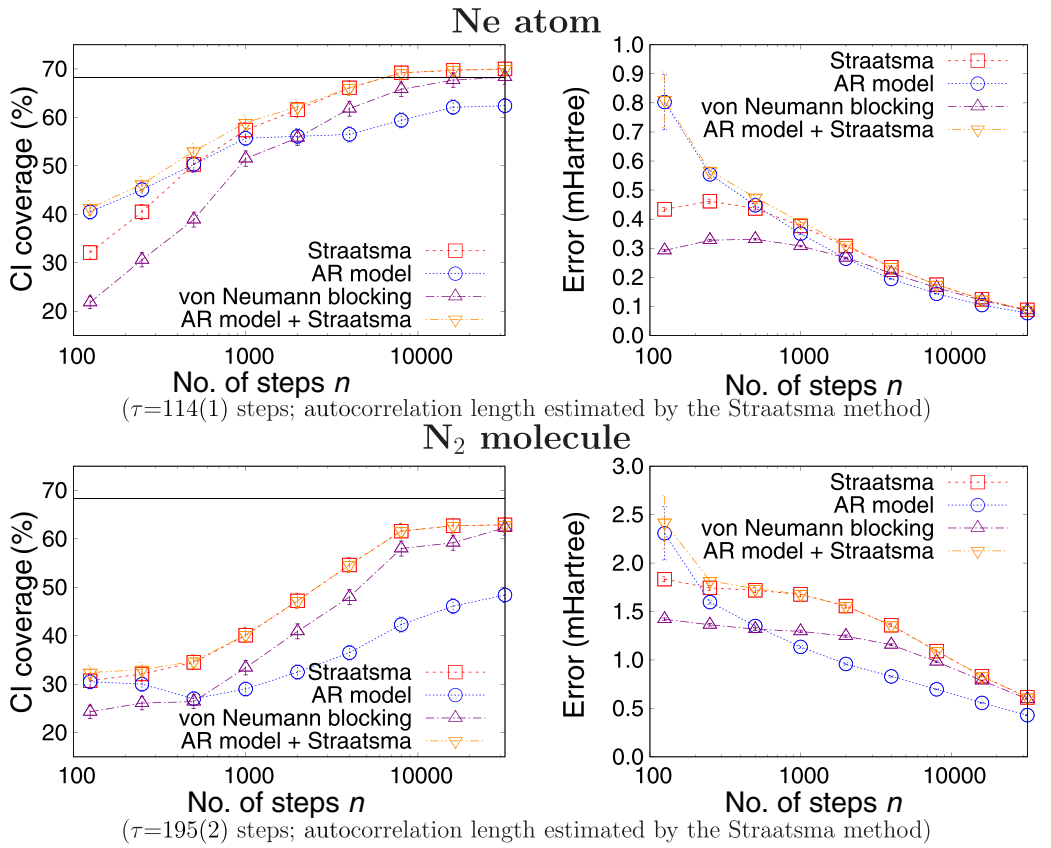

FIG. 3. Plots of the CI coverage and statistical error for the time series given by CCMC. The left and right columns show the estimated CI coverage and the statistical error of the energy mean obtained by applying the error estimation methods to the time series of the Ne atom and the N_2 molecule. The autocorrelation lengths τ are given by Straatsma.

time series for the Ne atom, the N_2 molecule, and the electron gas, respectively, using the DMC, FCIQMC, and CCMC methods. We discarded the beginning 10 000 steps to exclude the initialization bias. Then we applied the error estimation methods to them and surveyed the "confidence interval (CI) coverage" of 1σ confidence for different lengths of time series. Here, the CI coverage C is the fraction of the means of the 1000 time series which coincide with a reference mean value within the estimated errors. The condition for coincidence is given as

$$|\mu_i - \mu_{\text{reference}}| < \sigma_i. \tag{20}$$

Here, $\mu_{\text{reference}}$ is a reference mean (explained later in this paragraph), and μ_i and σ_i are the estimated mean and error of the mean for the ith time series. Let the number of time series that satisfy condition (20) be $n_{\text{coincidence}}$ out of 1000; the CI coverage is $C = n_{\text{coincidence}}/1000$. The error of C is evaluated by $\sqrt{C(1-C)/n_{\text{tot}}}$ (where n_{tot} is the number of time series), since the C is a mean of a binary variable (match or unmatch). The reference mean was obtained as an average over the 1000 time series. In the case of CCMC applied to the Ne atom, the difference between the reference mean and the deterministic CCSD value was just 6×10^{-6} Hartree. The expected C for 1σ confidence is 68.27% in theory. Therefore, when the observed C is closer to 68.27%, the error estimation is regarded to be more reliable.

We calculated the mean and error of the energy time series of DMC, FCIQMC, and CCMC by assuming that the weights of the samples are all equivalent, regardless of the differences of walker population. We verified that the bias caused by this approximation (i.e., difference between weighted and unweighted energy mean) is much smaller than the statistical error within the studied data size (see Appendix B). Therefore, we expect that the conclusions of this work will also apply if taking into account the difference of weights of each sample.

This analysis does not consider how much σ_i is underestimated for noncoincident cases. A possible way to investigate the outlier samples is plotting how far the mean is from the reference mean on a scale of the estimated error. However, for such an analysis, one has to prepare a very large number of time series to collect a sufficient number of outlier samples. Therefore, we evaluate the error estimation methods solely with the CI coverage in this work.

B. Comparison between DMC, FCIQMC and CCMC

Figures 2–4 show the CI coverage (left) and the statistical error of the mean (right) obtained by the error estimation methods. The initial 10 000 steps are removed beforehand to eliminate the equilibration phase. For every case, all the error estimation methods give lower CI coverage than the ideal value for the 1σ confidence interval, 68.27%, when the length of time series n is small. This is attributed to the

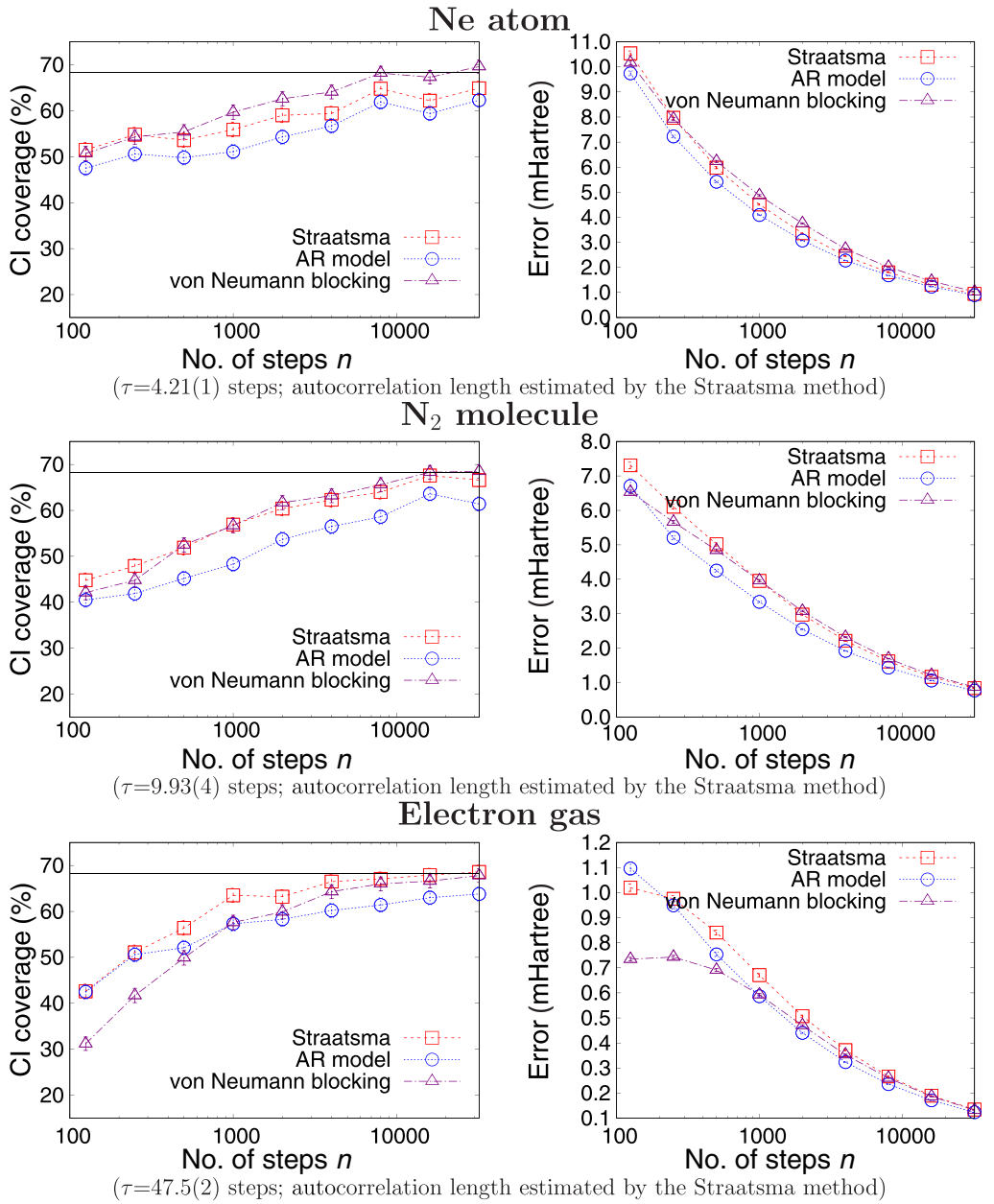

FIG. 4. Plots of the CI coverage and statistical error for the time series given by DMC. The left and right columns show the estimated CI coverage and statistical error of the energy mean given by applying the error estimation methods to the time series of the Ne atom, the N_2 molecule, and the electrons gas. The autocorrelation lengths τ are given by Straatsma.

underestimation of the error of the mean, which is typically observed in error estimation [24]. It is roughly estimated from the results in Appendix C how much the error of the mean would be underestimated for the given CI coverage.

First, we discuss the results of FCIQMC and CCMC. For comparatively short lengths of time series, the AR model gives the highest CI coverage. Moreover, comparison of the estimated statistical errors further distinguishes the AR model from the other methods: Only the AR model reproduces that the estimated errors normally decrease in proportion to $1/\sqrt{n}$ from the smallest number of steps. (For example, in the upper right figure of Fig. 2, the plot of the statistical error estimated by the AR model reasonably monotonically decreases with the number of steps n, while those by Straatsma and von Neumann blocking unreasonably increase with n when n is small.) This clearly shows the advantage of taking the AR model of Eq. (5). On the other hand, the lowest CI coverage is measured by von Neumann blocking. This is because the length of the time series is not large enough to prepare a significant number of blocks for the von Neumann's ratio test for randomness, in which case the time series is divided into 10 blocks and the statistical error is calculated with assuming that the 10 blocks are independent of each other, as mentioned in Sec. II C.

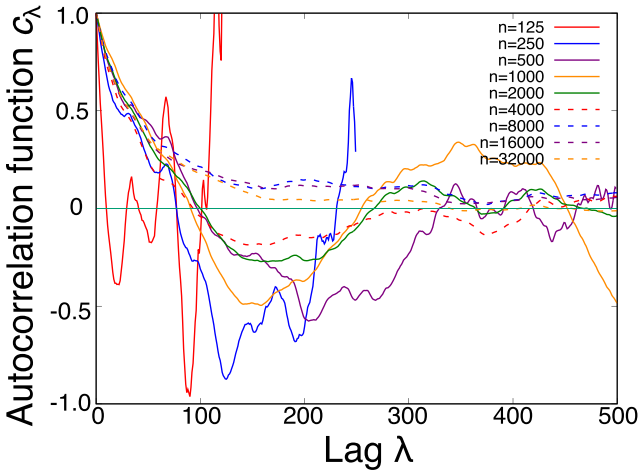

FIG. 5. Autocorrelation functions for different lengths n of time series, given by the CCMC method applied to the Ne atom. For comparatively small n, the autocorrelation function c_λ apparently includes a noise and it cannot be seen that c_λ gradually converges to zero along with the lag λ increasing.

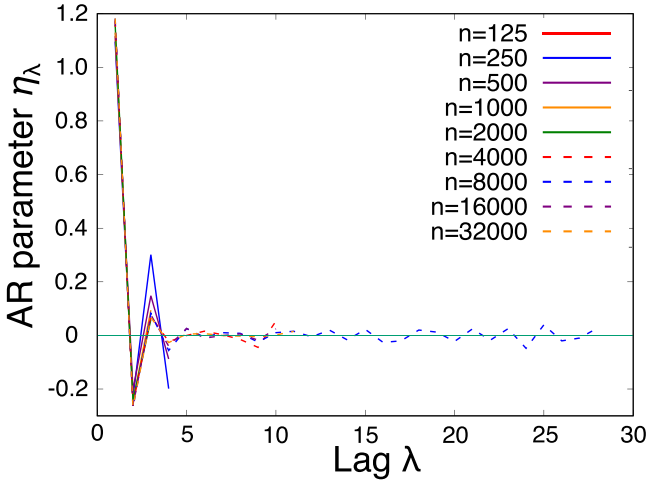

FIG. 6. Parameters η_λ in the AR model fitted to different lengths of time series given by the CCMC method applied to the Ne atom. The number of the parameters is determined by AIC [29]. The parameters are terminated or converged around zero within a few terms, regardless of the length of time series.

Straatsma gives intermediate CI coverages for small lengths n of the time series. Since the calculated τ fully depends on the autocorrelation function c_λ through Eq. (3), we examined how the shape of the autocorrelation function c_λ changes according to the length n of the time series in Fig. 5: The autocorrelation function c_λ becomes negative more quickly for smaller n by the c_λ oscillating since it is estimated by an insufficient number of terms, $n - \lambda$, through Eq. (4). The truncation of the sum in Eq. (3) is so drastic that Straatsma underestimates the autocorrelation time. In contrast, the oscillation of c_λ does not much affect the AR model, although its estimation also depends on autocorrelation functions c_λ through Eq. (8). This is because taking a product with η_λ drastically reduces the contribution of c_λ with large lag λ: Fig. 6 shows the expansion of the parameters η_λ, where they are terminated or converged to zero only within a few terms. Therefore, just a few terms of c_λ from small lag λ are used to calculate τ in the AR model.

When the length of time series n is comparatively large, the CI coverages of the AR model converges towards 68.27% the most slowly. This is because the assumption of Eq. (5) in the AR model cannot fully describe the target random process, and therefore its reliability is reduced. To summarize, the AR model (Straatsma) is the most reliable for small (middle and large) length time series, respectively. We have therefore devised a hybrid scheme of the AR model and Straatsma, which works reasonably for any length of time series. It simply adopts the larger of the errors estimated by both methods: $\sigma_{\bar{X}}(\text{hybrid}) = \max\{\sigma_{\bar{X}}(\text{AR model}), \sigma_{\bar{X}}(\text{Straatsma})\}$. The CI coverages for the hybrid method are shown as the "AR model + Straatsma" in Figs. 2 and 3 and are always comparatively the closest to 68.27%.

We similarly evaluate the error estimation methods for DMC time series. The AR model gives the lowest CI coverages for a Ne atom and N_2 molecule (see Fig. 4), unlike the case of FCIQMC and CCMC (see Figs. 2 and 3). The different performance of the AR model between DMC, FCIQMC, and CCMC may be attributed to a different degree of leptokurtosis between the DMC's, FCIQMC, and CCMC's sampling distributions. The AR model assumes that a noise of each sample follows a normal distribution. Therefore, when the distribution of actual noise is far from a normal distribution, the performance of the AR model would decay. The sampling distribution of DMC is known to be leptokurtotic [42–44]. Meanwhile, a sampling distribution of FCIQMC and CCMC is generally not leptokurtotic because the energy estimator given in Eq. (19) is bounded. The ingredients of the Hamiltonian matrix, $\langle D_0|\hat{H}|D_0\rangle$, $\langle D_0|\hat{H}|D_i^a\rangle$, and $\langle D_0|\hat{H}|D_{i,j}^{a,b}\rangle$, are always all bounded. c_i^a/c_0 and $c_{i,j}^{a,b}/c_0$ are less than one if the multireference character of the system is weak. When FCIQMC is applied to sufficiently multireference systems, c_i^a/c_0 and $c_{i,j}^{a,b}/c_0$ might be extremely large since c_0 possibly becomes very small. In such cases, coupled cluster theory will generally not converge, and the CCMC algorithm will fail notably. However, our target systems are not such multireference systems. Therefore, we conclude that the leptokurtosis that is unique for the DMC's sampling distributions in this study leads to the lower performance of the AR model in DMC. Appendix D shows that the distribution of "skewness" and "kurtosis" of the DMC time series has the characteristic dependence of a leptokurtotic distribution.

Among the DMC results, in the case of electron gas, the AR model's CI coverage is as high as the Straatsma method for smaller lengths of time series (see the bottom figures in Fig. 4). This is attributed to the sampling distribution of electron gas that is expected to have less heavy tails. The heavy tails are caused by cusps in the wave function [42–45]. Since electron gas does not contain atomic cores, electron-core cusps are not in the wave function. Since the AR model is the most affected by heavy tails, it makes sense that the absence of the electron-core cusp in electron gas improves the AR model's performance compared with the cases of the Ne atom and N_2 molecule.

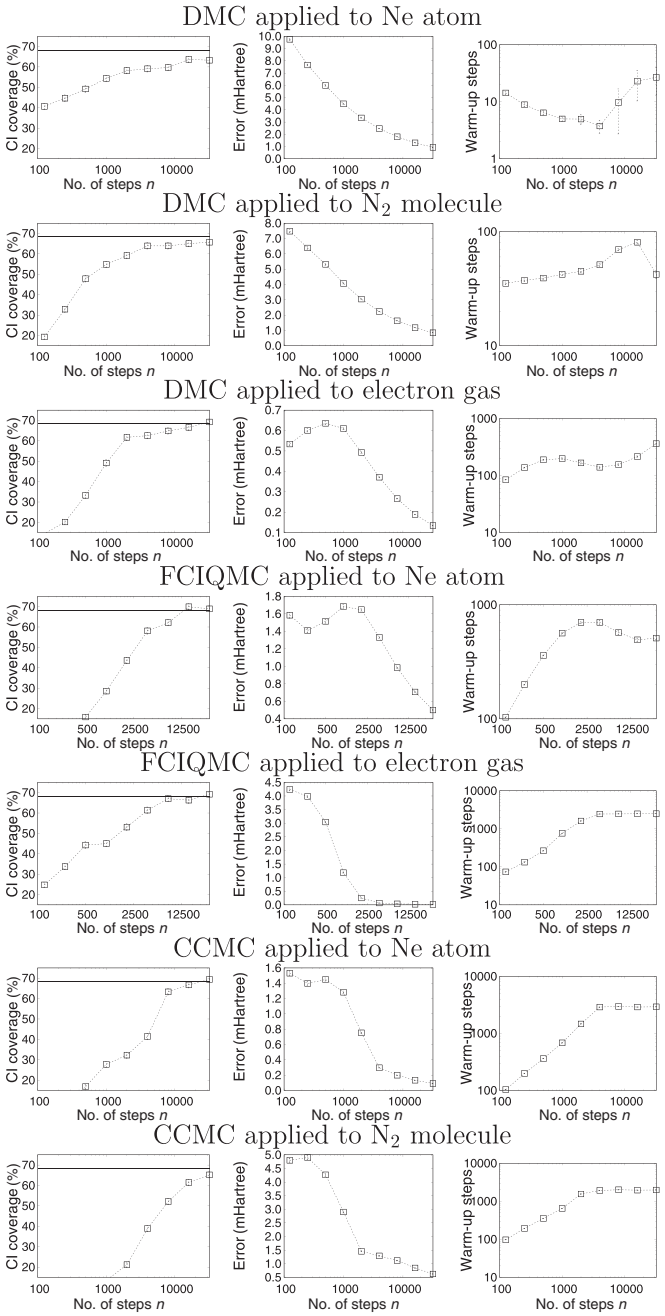

FIG. 7. The right column figures show the estimated number of warm-up steps computed by the warm-up steps estimation scheme, the MSER method. The left two columns show the CI coverage and the error of the mean given by Straatsma applied to the time series after eliminating the estimated warm-up steps. The horizontal axis value is the length of the time series before the elimination.

τ is written in Figs. 2–4 representing the autocorrelation time given by the Straatsma method. For the Ne atom and N_2 molecule, the autocorrelation time is much smaller for DMC than for FCIQMC and CCMC. A smaller autocorrelation time indicates that there is an effectively larger number of independent samples in the time series. Therefore, the CI coverage is expected to increase more rapidly for DMC than for FCIQMC and CCMC. However, the plot of CI coverage of Straatsma is not as different as the difference of the autocorrelation time between DMC, FCIQMC, and CCMC. Therefore, we conclude that the Straatsma method tends to underestimate τ more severely for DMC than FCIQMC and CCMC. This would be attributed to heavy tails existing in a DMC sampling distribution, unlike FCIQMC and CCMC. Heavy tails would increase the variance of autocorrelation functions c_λ defined in Eq. (4). Such a larger noise in c_λ causes an earlier truncation of integration in Eq. (3) before c_λ reaches a plateau, by satisfying the truncation condition $c_\lambda < 0$. This should cause an underestimate of the autocorrelation time τ. We introduce different methods to evaluate the integration of autocorrelation functions in Appendix E. We also discuss the relationship between the number of correlation steps and the QMC time steps in Appendix F.

C. MSER: Estimation of number of warm-up steps

We discuss the performance of the MSER method for the warm-up steps estimation. We applied the MSER method to the full time series of the energy estimator, including the nonplateau part, and removed the estimated warm-up steps d. Then, we applied the Straatsma method to obtain the CI coverage and the statistical error. Figure 7 shows the CI coverage (left), the statistical error (middle), and the estimated warm-up steps (right). The plot of the number of estimated warm-up steps reaches a plateau at a certain length of time series, which is a reasonable behavior. In addition, the CI coverage is mostly around 68.26% when the length of time series is as large as the number of the estimated warm-up steps is in the plateau part: The plateau value is sufficiently large to remove the initialization bias.

As an extension to this study, the performance of different schemes discussed in the literature [46–49] could be compared to MSER. Chodera [49], for example, determined the starting iteration to be the iteration with the most "independent data points" (compare roughly to k in blocking analysis). Another example is the starting point finder described in Yang *et al.* [46], who first divided up the data set into independent data points and then, starting from the end, analyzed whether these data points are still normally distributed as they kept including earlier points in their test—the point when they no longer were normally distributed served as the starting point estimate.

VI. CONCLUSION

A QMC time series consists of (a) the equilibration region and (b) the plateau region. An estimated energy of QMC is given as the mean of region (b) that has a statistical error. One needs to estimate the error with a reliable analysis method. We evaluated the performance of three characteristic methods for error estimation: The AR model, Straatsma method, and von Neumann blocking, with an analysis based on CI coverage. The comparison among the three methods showed a qualitatively different trend between that for DMC and those for FCIQMC and CCMC. For the DMC case, the von Neumann blocking was predicted to be the most reliable. This is attributed to the existence of heavy tails in DMC's sampling distribution: While the applicabilities of the AR model and Straatsma are restricted within the scope of their

assumption, such as the normality in the target distributions or some specific modeling on the autocorrelation length, von Neumann blocking does not rely on such stronger assumption relatively, being robust to heavy tails. For DMC, it is therefore safer to apply von Neumann blocking for the analysis. For the FCIQMC and CCMC cases, since the heavy tails never exist, the Straatsma method and AR model became reliable. We found that the AR model (the Straatsma method) was the most reliable for shorter (longer) length of time series. We further developed a different approach named the hybrid method, which takes the advantages of the Straatsma method and AR model. The hybrid method was the most reliable for any length of time series.

It is also important to reliably estimate the length of the region (a) to exclude it prior to estimating an error of the mean in the region (b). We established that the MSER method [27,30] gave a reasonable estimation of the number of warm-up steps.

Research data supporting this work and further information can be found at [50].

ACKNOWLEDGMENTS

We used GNUPLOT to make figures [51] and MATPLOTLIB [52] of PYTHON 3 [53]. For our analyzes, we used PYTHON 3 [53] and the libraries numpy [54], scipy [55], and pandas [56]. The computation in this work has been performed using the facilities of the Research Center for Advanced Computing Infrastructure (RCACI) at JAIST. We thank the referees for many significant comments about the skewness and kurtosis of the energy estimator of DMC, relationship between walker move denial rate and correlation time, leptokurtosis of FCIQMC and CCMC, etc. T.I. is grateful for financial support from a Grant-in-Aid for JSPS Research Fellow (Grant No. 18J12653). V.A.N. thanks Samuel Greene for helpful discussions. V.A.N. would like to acknowledge the EPSRC Centre for Doctoral Training in Computational Methods for Materials Science for funding under Grant No. EP/L015552/1. K.H. is grateful for financial support from the HPCI System Research Project (Project ID No. hp190169), MEXT-KAKENHI (Grants No. JP16H06439, No. JP17K17762, No. JP19K05029, and No. JP19H05169), and the Air Force Office of Scientific Research (Award No. FA2386-20-1-4036). R.M. is grateful for financial support from MEXT-KAKENHI (Grants No. 17H05478 and No. 16KK0097), Toyota Motor Corporation, I-O DATA Foundation, and the Air Force Office of Scientific Research (Grant No. AFOSR-AOARD/FA2386-17-1-4049). R.M. and K.H. are also grateful for financial support from MEXT-FLAGSHIP2020 (Project ID No. hp170269 and No. hp170220). A.J.W.T. thanks the Royal Society for a University Research Fellowship (Grant No. UF110161).

APPENDIX A: EVALUATION OF autoCORRELATION LENGTH τ BY EQ. (3)

The Straatsma method estimates the autocorrelation length τ by accumulating the autocorrelation functions c_λ in Eq. (3). Since the approximation of c_λ by Eq. (4) is less accurate for larger λ, the accumulation should be truncated at an appropriate small λ. For the results in the main text, we took a summation until c_λ became negative for the first time (this truncation point is denoted as λ^*). Here, we verified how reasonable this criterion is by checking how the CI coverage changes depending on the number of autocorrelation functions to be summed up: $\lambda^* \times$ 0.50, 0.75, 1.00, 1.25, and 1.50. The result is shown in Fig. 8. 0.50 λ^* gives significantly lower

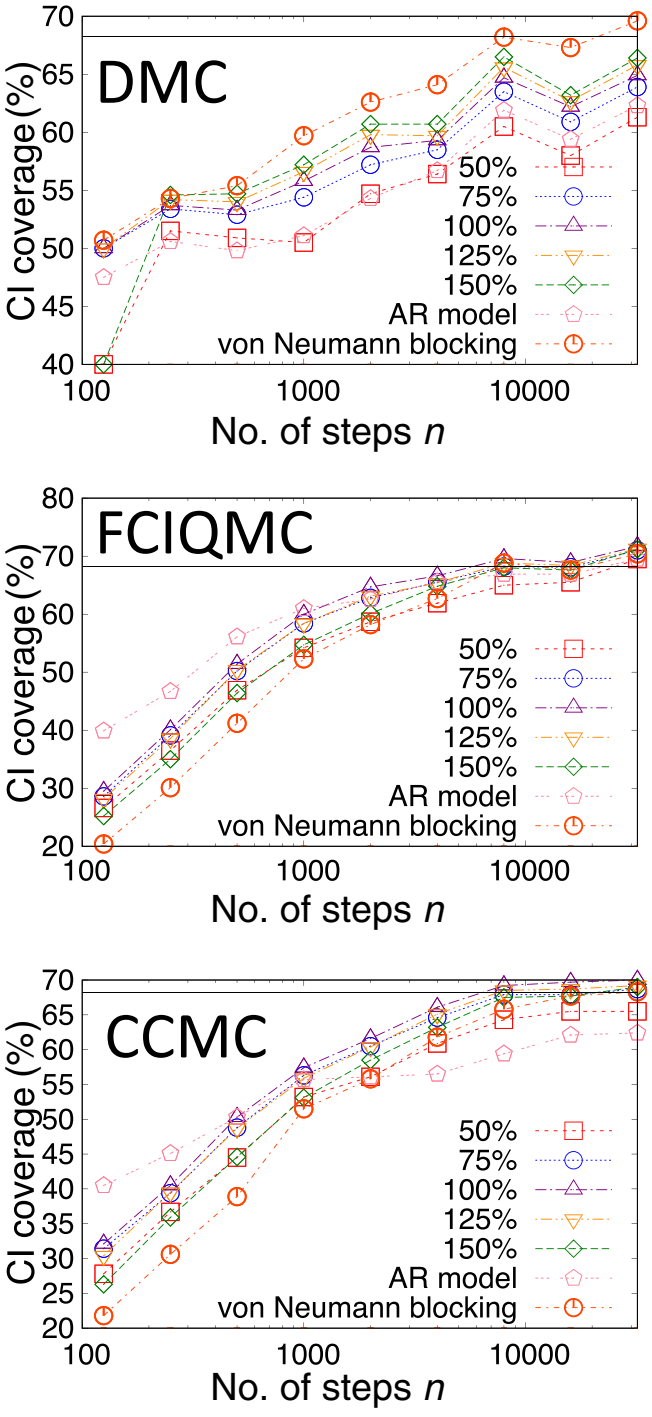

FIG. 8. CI coverages obtained by the Straatsma method with different truncation points of Eq. (3). We applied this method to the time series of the Ne atom computed by DMC, FCIQMC, and CCMC, respectively.

TABLE I. Mean of $r_{\text{bias/error}}$ and unbiased estimation of the standard deviation of $r_{\text{bias/error}}$'s distribution.

Method	System	Mean of $r_{\text{bias/error}}$	Std. dev. of $r_{\text{bias/error}}$
DMC	Ne atom	0.056	0.027
	N_2 molecule	0.083	0.019
	Electron gas	0.255	0.034
FCIQMC	Ne atom	0.079	0.014
	Electron gas	0.081	0.021
CCMC	Ne atom	0.046	0.006
	N_2 molecule	0.014	0.004

CI coverage than the ideal value of 68.27% for every QMC result, since it underestimates the autocorrelation length by ignoring some c_λ before reaching the plateau. In the other plots of CI coverage, 1.00, 1.25, and 1.50 λ^* are quite similar and at least the difference is smaller than the difference by error estimations methods (i.e., the Straatsma method, the AR model, and the von Neumann blocking). Therefore the choice of the truncation point would not qualitatively change our discussion.

APPENDIX B: BIAS IN THE ENERGY MEAN DUE TO IGNORING WALKER POPULATION DIFFERENCES

We calculated the ratio $r_{\text{bias/error}}$ of bias in the energy mean attributed to ignoring the differences of the walker population,

$$\left| \frac{1}{n} \sum_{i=1}^{n} e_i - \frac{\sum_{i=1}^{n} w_i e_i}{\sum_{i=1}^{n} w_i} \right|, \tag{B1}$$

and the error in the energy mean given by the Straatsma method without considering the differences of population. Here, e_i is the ith energy sample and w_i is the corresponding weight, for the total population of walkers. Although it is not proved that the total number of walkers is regarded as the weight of the instantaneous projected energy, we think this is, at least, a reasonable approximation to estimate the bias of ignoring population differences. We calculated $r_{\text{bias/error}}$ for the 1000 times series with 32 000 steps, which is the maximum data size studied in the paper. Table I shows the mean of $r_{\text{bias/error}}$ and unbiased estimation of the standard deviation of $r_{\text{bias/error}}$'s distribution. The largest $r_{\text{bias/error}}$ was obtained for the case that DMC is applied to the electron gas, but the bias is much smaller than the error.

APPENDIX C: RELATIONSHIP BETWEEN THE CI COVERAGE AND UNDERESTIMATION OF THE STATISTICAL ERROR

We show in this paper that the CI coverage tends to be smaller than ∼68.27% (expected value for 1σ error bars) for shorter lengths of the time series. This is attributed to the underestimation of the statistical error of the mean. Here, we would like to directly show how much the error of the mean is underestimated for the given CI coverage. For this purpose, we compared CI coverage and $\sigma_{\overline{X}}(n) \cdot \sqrt{n}$. Here, n is the number of samples in the time series and $\sigma_{\overline{X}}(n)$ is the error of the mean calculated for the time series with n samples.

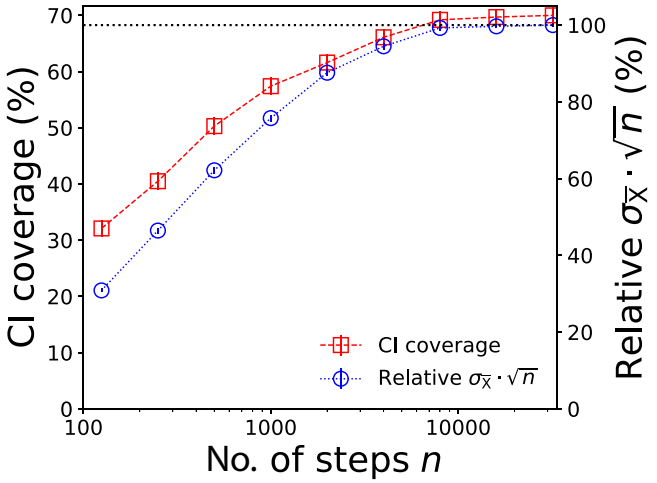

FIG. 9. Comparison of the CI coverage and relative $\sigma_{\overline{X}}(n) \cdot \sqrt{n}$. The CI coverage and statistical error were calculated by the Straatsma method for the time-series of CCMC applied to Ne atom.

The equilibration steps are removed from the time series in advance. $\sigma_{\overline{X}}(n) \cdot \sqrt{n}$ does not depend on n, in principle, but it will be smaller if $\sigma_{\overline{X}}(n)$ is underestimated. Therefore, a comparison of $\sigma_{\overline{X}}(n) \cdot \sqrt{n}$ between smaller n values and a sufficiently large n indicates how much $\sigma_{\overline{X}}(n)$ with smaller n values are underestimated. Figure 9 shows the relationship between CI coverage and $\sigma_{\overline{X}}(n) \cdot \sqrt{n}$ calculated for the results of the Straatsma method applied to CCMC for Ne. Here, $\sigma_{\overline{X}}(n) \cdot \sqrt{n}$ is given as a percentage with the value at $n = 32\,000$ being 100%. To extract some values, for $n = 125$, the CI coverage is 32.1(1.5)% and the relative $\sigma_{\overline{X}}(n) \cdot \sqrt{n}$ is 30.9(5)%. For $n = 1000$, the CI coverage was 57.4(1.6)%, while the relative $\sigma_{\overline{X}}(n) \cdot \sqrt{n}$ was 75.8(9)%.

APPENDIX D: SKEWNESS AND KURTOSIS

The skewness and kurtosis of the local energy estimator of DMC are not generally statistically meaningful [42–44] because the sampling distribution of the energy estimator is leptokurtotic and the "skewness" and "kurtosis" are dominated by the most outlying sample ("outlier"). Let the most outlier sample be E_{outlier} and the mean of the samples be E_{mean}; then the skewness and kurtosis are approximated,

$$\text{(Estimated skewness)} \sim (E_{\text{outlier}} - E_{\text{mean}})^3,$$
$$\text{(Estimated kurtosis)} \sim (E_{\text{outlier}} - E_{\text{mean}})^4.$$

Therefore, the skewness and kurtosis are related by

$$\text{(Estimated skewness)} \sim (\text{Estimated kurtosis})^{4/3}.$$

Figure 10 actually shows that the skewness and kurtosis calculated for the Ne atom are dependent for the DMC case, while the skewness and kurtosis are independent for the FCIQMC case.

APPENDIX E: OTHER EVALUATION METHODS OF THE AUTOCORRELATION LENGTH THAN STRAATSMA

We introduce some methods that estimate the integration of autocorrelation functions differently. Wolff's Γ method

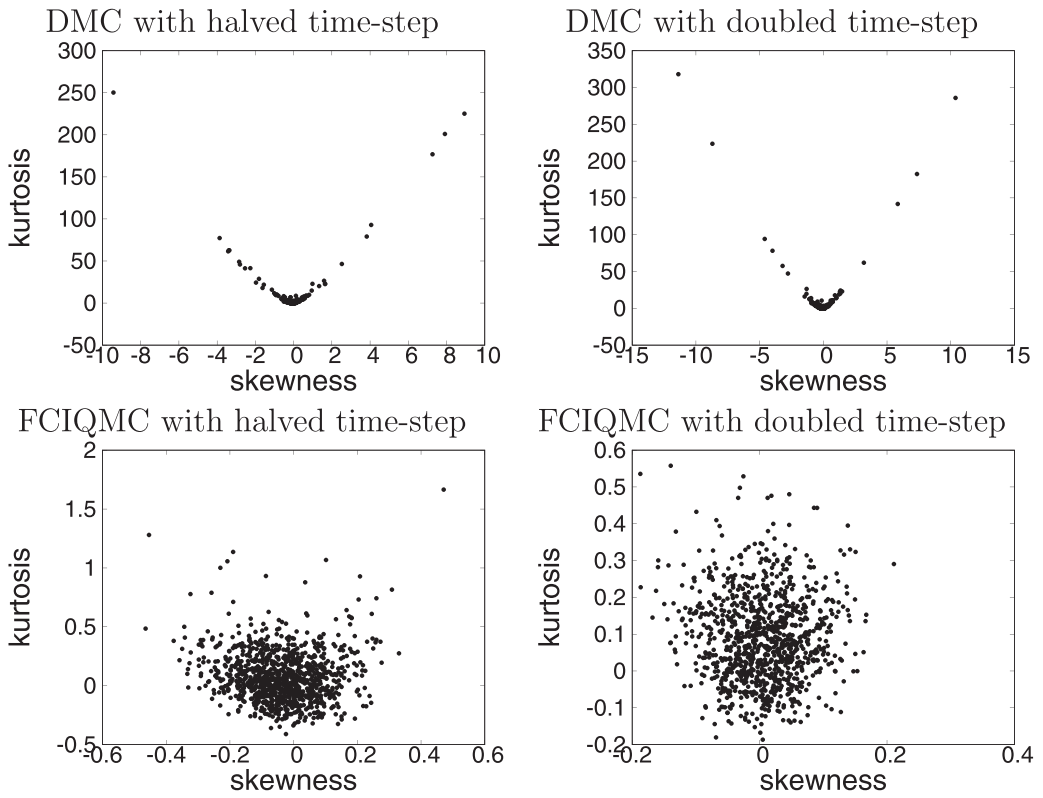

FIG. 10. The skewness and kurtosis computed for the time series given by DMC and FCIQMC applied to the Ne atom with halved and doubled time steps.

employs the following assumption. Let the autocorrelation time given by a very long length of time series be τ_{true} and an autocorrelation time calculated by integrating autocorrelation functions c_λ from $\lambda = 1$ to $\lambda_{\max}$ be $\tau(\lambda_{\max})$; then the bias in $\tau(\lambda_{\max})$ divided by τ_{true} is approximated by

$$\left|\frac{\tau(\lambda_{\max}) - \tau_{\text{true}}}{\tau_{\text{true}}}\right| \sim \exp\left(-\frac{\lambda_{\max}}{\tau}\right). \tag{E1}$$

They also show that the statistical error of $\tau(\lambda_{\max})$, with the mean of $\tau(\lambda_{\max})$ assumed τ_{true}, divided by τ_{true} is approximated by

$$\frac{\text{Err}[\tau(\lambda_{\max})]}{\tau_{\text{true}}} \sim \sqrt{\frac{2(2\lambda_{\max}+1)}{n}}. \tag{E2}$$

Here, n is the length of time series. The Γ method decides the $\lambda_{\max}$ to minimize a sum of relative bias in Eq. (E1) and error in Eq. (E2).

Another method is a blocking method by Flyvbjerg and Petersen [57]. This method applies a blocking operation to the time series with increasing the block size, $m = 1 \to 2 \to 4 \to \cdots$, and calculates the variance of the mean of the blocks with assuming that the blocks are independent of each other for every m. For smaller m, the variance increases with m increasing. Once m becomes large, as the blocks are independent of each other (i.e., the autocorrelation function $c_{\lambda=1}$ of the blocks equals zero), the variance oscillates around a constant value. The plateau values correspond to the variance of the mean. Since the error in the variance is larger for larger m, one should select m at the beginning of the plateau to yield the variance. However, this blocking analysis does not provide a way to automatically decide m. A recent work by Jonsson [58], based on a similar concept to Flyvbjerg and Petersen, provides an automatic algorithm to decide the block size m. The method selects the minimum m satisfying the null hypothesis that for all the block sizes larger than m, the autocorrelation functions (lag $\lambda = 1$) of the time series of block means are zero.

APPENDIX F: RELATIONSHIP BETWEEN THE CORRELATION LENGTH AND QMC TIME STEPS

We investigate the relationship between the number of correlated steps and the QMC time step. At first thought, the relationship should be inversely proportional. Figure 11 shows how the number of correlated steps estimated by Straatsma changes depending on the QMC time step for the time series of DMC for a Ne atom and FCIQMC for a Ne atom. The autocorrelation lengths of halved and doubled time steps are shown as ratios from that of the original time step. In the case of FCIQMC, the autocorrelation length is fully inversely proportional to the time step. Meanwhile, in the case of DMC, the relationship is not fully inversely proportional. A possible explanation for this would be the limitation scheme for the drift velocity and local energy as applied in the practical calculations [34]. The scheme is introduced to suppress the divergence of the drift velocity and local energy occurring near the nodal surfaces, where the walkers update is adjusted from that with constant time step dt, and hence the scheme modifies the naively expected dt dependence as $\sim 1/dt$. Another pos-

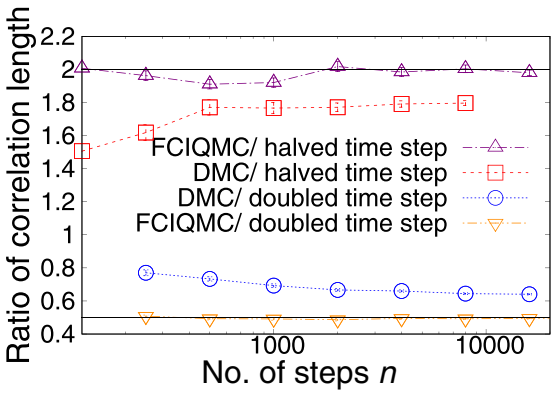

FIG. 11. The ratio between the autocorrelation lengths estimated for the time series with the original and halved (doubled) time steps. The time series are given by applying DMC and FCIQMC methods to the Ne atom. The Straatsma method is applied to calculate the autocorrelation lengths. The upper (lower) horizontal line indicates 2 (0.5).

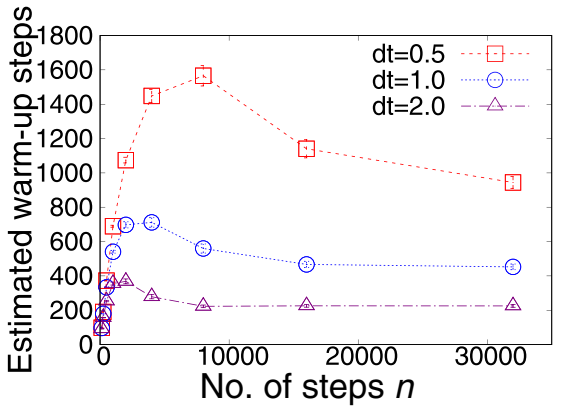

FIG. 12. The number of warm-up steps estimated by the MSER method for original, halved, and doubled time steps. This estimation is performed for the time series given by applying FCIQMC to the Ne atom.

sibility is the $O(dt^2)$ error in the DMC Green's function [9]. This also introduces another form of dt dependence, invading the naively expected $\sim 1/dt$ behavior. As such, the clear behavior of dt dependence for the autocorrelation does not hold, and hence one cannot precisely interpolate the autocorrelation time from some selected dt samples.

Figure 12 shows a relationship between the number of warm-up steps and the time step for FCIQMC for a Ne atom. We observe the expected inversely proportional relationship, as well as the number of autocorrelation steps. We cannot investigate the relationship for DMC because the equilibration part is tiny and the relative error of the estimated number of warm-up steps is too large to perform this analysis.

[1] J. S. Spencer, V. A. Neufeld, W. A. Vigor, R. S. T. Franklin, and A. J. W. Thom, Large scale parallelization in stochastic coupled cluster, J. Chem. Phys. **149**, 204103 (2018).

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

[3] A. Mathuriya, Y. Luo, R. C. Clay, III, A. Benali, L. Shulenburger, and J. Kim, Embracing a new era of highly efficient and productive quantum Monte Carlo simulations, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '17, edited by B. Mohr and P. Raghavan (ACM, New York, 2017), pp. 38:1–38:12.

[4] T. Ichibha, Z. Hou, K. Hongo, and R. Maezono, New insight into the ground state of fepc: A diffusion Monte Carlo study, Sci. Rep. **7**, 2011 (2017).

[5] Y. Luo, A. Benali, L. Shulenburger, J. T. Krogel, O. Heinonen, and P. R. C. Kent, Phase stability of tio$_2$ polymorphs from diffusion quantum Monte Carlo, New J. Phys. **18**, 113049 (2016).

[6] J. Trail, B. Monserrat, P. López Ríos, R. Maezono, and R. J. Needs, Quantum Monte Carlo study of the energetics of the rutile, anatase, brookite, and columbite tio$_2$ polymorphs, Phys. Rev. B **95**, 121108(R) (2017).

[7] Q. S. Ken, T. Ichibha, K. Hongo, and R. Maezono, Difficulty to capture non-additive enhancement of stacking energy by conventional *ab initio* methods, Chem. Phys. **529**, 110554 (2020).

[8] A. Benali, Y. Luo, H. Shin, D. Pahls, and O. Heinonen, Quantum Monte Carlo calculations of catalytic energy barriers in a metallorganic framework with transition-metal-functionalized nodes, J. Phys. Chem. C **122**, 16683 (2018).

[9] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, Rev. Mod. Phys. **73**, 33 (2001).

[10] J. B. Anderson, A random-walk simulation of the Schrödinger equation: H_3^+, J. Chem. Phys. **63**, 1499 (1975).

[11] J. B. Anderson, Quantum chemistry by random walk. h^2 p, h+3 d3h 1á1, h2 3Σ+u, h4 1Σ+g, be 1s, J. Chem. Phys. **65**, 4121 (1976).

[12] G. Ortiz, D. M. Ceperley, and R. M. Martin, New Stochastic Method for Systems with Broken Time-Reversal Symmetry: 2d Fermions in a Magnetic Field, Phys. Rev. Lett. **71**, 2777 (1993).

[13] P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, Inhomogeneous backflow transformations in quantum Monte Carlo calculations, Phys. Rev. E **74**, 066701 (2006).

[14] M. A. Morales, J. McMinis, B. K. Clark, J. Kim, and G. E. Scuseria, Multideterminant wave functions in quantum Monte Carlo, J. Chem. Theory Comput. **8**, 2181 (2012).

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

[16] G. H. Booth and A. Alavi, Approaching chemical accuracy using full configuration-interaction quantum Monte Carlo: A study of ionization potentials, J. Chem. Phys. **132**, 174104 (2010).

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

[18] D. M. Cleland, G. H. Booth, and A. Alavi, A study of electron affinities using the initiator approach to full configuration interaction quantum Monte Carlo, J. Chem. Phys. **134**, 024112 (2011).

[19] L. Veis, A. Antalík, Ö. Legeza, A. Alavi, and J. Pittner, The intricate case of tetramethyleneethane: A full configuration interaction quantum Monte Carlo benchmark and multireference coupled cluster studies, J. Chem. Theory Comput. **14**, 2439 (2018).

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

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

[22] C. Alexopoulos, D. Goldsman, P. Tang, and J. R. Wilson, Spsts: A sequential procedure for estimating the steady-state mean using standardized time series, IIE Trans. **48**, 864 (2016).

[23] T. Straatsma, H. Berendsen, and A. Stam, Estimation of statistical errors in molecular simulation calculations, Mol. Phys. **57**, 89 (1986).

[24] M. B. Thompson, Graphical comparison of mcmc performance, arXiv:1011.4457.

[25] G. S. Fishman, *Principles of Discrete Event Simulation* (Wiley, New York, 1978).

[26] S. Yousefi, MSER-5Y: An improved version of MSER-5 with automatic confidence interval estimation, Master's thesis, North Carolina State University, 2011.

[27] W. W. Franklin and K. P. White, Stationarity tests and mser-5: Exploring the intuition behind mean-squared-error-reduction in detecting and correcting initialization bias, in *2008 Winter Simulation Conference* (IEEE, 2008), pp. 541–546.

[28] W. Wei, *Time Series Analysis: Univariate and Multivariate Methods* (Springer, New York, 2006).

[29] H. Akaike, Information theory and an extension of the maximum likelihood principle, in *Breakthroughs in Statistics: Foundations and Basic Theory*, edited by S. Kotz and N. L. Johnson (Springer, New York, 1992), pp. 610–624.

[30] J. K. Preston White, An effective truncation heuristic for bias reduction in simulation output, Simulation **69**, 323 (1997).

[31] R. J. Needs, M. D. Towler, N. D. Drummond, P. López Ríos, and J. R. Trail, Variational and diffusion quantum Monte Carlo calculations with the CASINO code, J. Chem. Phys. **152**, 154106 (2020).

[32] V. Saunders, R. Dovesi, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, N. Harrison, K. Doll, B. Civalleri, I. Bush, P. D'Arco, and M. Llunell, *CRYSTAL03 User's Manual* (University of Torino, Torino, 2003).

[33] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.105.045313 for CRYSTAL03 input files to generate trial wave functions of DMC calculations for a neon atom and nitrogen molecule.

[34] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, A diffusion Monte Carlo algorithm with very small time-step errors, J. Chem. Phys. **99**, 2865 (1993).

[35] A. Ma, M. D. Towler, N. D. Drummond, and R. J. Needs, Scheme for adding electron-nucleus cusps to Gaussian orbitals, J. Chem. Phys. **122**, 224322 (2005).

[36] T. Kato, On the eigenfunctions of many-particle systems in quantum mechanics, Commun. Pure Appl. Math. **10**, 151 (1957).

[37] J. S. Spencer, N. S. Blunt, W. A. Vigor, F. D. Malone, W. M. C. Foulkes, J. J. Shepherd, and A. J. W. Thom, Open-source development experiences in scientific software: The Hande quantum Monte Carlo project, J. Open Source Softw. **3**, e9 (2015).

[38] J. S. Spencer, N. S. Blunt, S. Choi, J. Etrych, M.-A. Filip, W. M. C. Foulkes, R. S. T. Franklin, W. J. Handley, F. D. Malone, V. A. Neufeld, R. Di Remigio, T. W. Rogers, C. J. C. Scott, J. J. Shepherd, W. A. Vigor, J. Weston, R. Xu, and A. J. Thom, The Hande-QMC project: Open-source stochastic quantum chemistry from the ground state up, J. Chem. Theory Comput. **15**, 1728 (2019).

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

[40] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. u. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. r. m. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability, J. Chem. Theory Comput. **13**, 3185 (2017).

[41] U. Wolff, Monte Carlo errors with less errors, Comput. Phys. Commun. **156**, 143 (2004).

[42] J. R. Trail, Heavy-tailed random error in quantum Monte Carlo, Phys. Rev. E **77**, 016703 (2008).

[43] J. R. Trail and R. Maezono, Optimum and efficient sampling for variational quantum Monte Carlo, J. Chem. Phys. **133**, 174120 (2010).

[44] P. L. Ríos and G. J. Conduit, Tail-regression estimator for heavy-tailed distributions of known tail indices and its application to continuum quantum Monte Carlo data, Phys. Rev. E **99**, 063312 (2019).

[45] J. Tiihonen, R. C. Clay, and J. T. Krogel, Toward quantum Monte Carlo forces on heavier ions: Scaling properties, J. Chem. Phys. **154**, 204111 (2021).

[46] W. Yang, R. Bitetti-Putzer, and M. Karplus, Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence, J. Chem. Phys. **120**, 2618 (2004).

[47] S. K. Schiferl and D. C. Wallace, Statistical errors in molecular dynamics averages, J. Chem. Phys. **83**, 5203 (1985).

[48] J. T. Krogel, Nexus: A modular workflow management system for quantum simulation codes, Comput. Phys. Commun. **198**, 154 (2016).

[49] J. D. Chodera, A simple method for automated equilibration detection in molecular simulations, J. Chem. Theory Comput. **12**, 1799 (2016).

[50] https://doi.org/10.17863/CAM.83172.

[51] T. Williams, C. Kelley *et al.*, Gnuplot 5.4: An interactive plotting program, http://gnuplot.sourceforge.net/.

[52] J. D. Hunter, MATPLOTLIB: A 2d graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[53] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).

[54] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, Array programming with NUMPY, Nature (London) **585**, 357 (2020).

[55] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, SciPy 1.0: Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).

[56] W. McKinney, Data Structures for Statistical Computing in Python, in *Proceedings of the 9th Python in Science Conference*, edited by S. van der Walt and J. Millman (2010), pp. 56–61.

[57] H. Flyvbjerg and H. G. Petersen, Error estimates on averages of correlated data, J. Chem. Phys. **91**, 461 (1989).

[58] M. Jonsson, Standard error estimation by an automated blocking method, Phys. Rev. E **98**, 043304 (2018).