

# Reducing autocorrelation time in determinant quantum Monte Carlo using Wang-Landau algorithm: application to Holstein model

Meng Yao,<sup>1</sup> Da Wang,<sup>1,2,\*</sup> and Qiang-Hua Wang<sup>1,2,†</sup>

<sup>1</sup>*National Laboratory of Solid State Microstructures & School of Physics, Nanjing University, Nanjing, 210093, China*

<sup>2</sup>*Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing 210093, China*

When performing a Monte Carlo calculation, the running time should in principle be much longer than the autocorrelation time in order to get reliable results. Among different lattice fermion models, the Holstein model is notorious for its particularly long autocorrelation time. In this work, we employ the Wang-Landau algorithm in the determinant quantum Monte Carlo to achieve the flat-histogram sampling in the “configuration weight space”, which can greatly reduce the autocorrelation time by sacrificing some sampling efficiency. The proposal is checked in the Holstein model on both square and honeycomb lattices. Based on such a Wang-Landau assisted determinant quantum Monte Carlo method, some models with long autocorrelation times can now be simulated possibly.

When applying quantum Monte Carlo (MC) to correlated fermion systems, the negative sign problem is in general found to be nondeterministic polynomial hard [1] as the sign average  $\langle s \rangle \propto e^{-a\beta N}$  approaches zero exponentially versus the imaginary time  $\beta$  (inverse of temperature) times particle number  $N$ , which inevitably demands exponential computational effort to get controlled data when approaching thermodynamic limit. Similarly, the same thing happens to the problems with exponential autocorrelation time even without sign problem. Moreover, if the autocorrelation time  $\tau_a$  is sufficiently long, we cannot simply divide the whole Markov chain into several short ones (walkers) to reduce the statistical error unless the starting point of each walker can be chosen uniformly at random on the whole Markov chain [2], which, however, is difficult to realize due to the nonuniform density of states (DOS) of the “configuration weight space” (corresponding to energy space in classical models).

When designing a MC algorithm, it is very important to make the autocorrelation time short enough. [3] To this end, roughly two classes of methods have been broadly used: cluster algorithms [4, 5] and reweighting techniques [6–8]. Most of them work pretty well for discrete classical models such as Ising and Potts models, but for a quantum model, it is difficult to design a powerful method in general. Global update in principle can greatly reduce the autocorrelation time but always at the price of severely losing the algorithm efficiency or yielding very low acceptance rate. [9] On the other hand, reweighting by another parameter set [6] or other replicas [8] are often used to avoid the ergodicity breaking. In special, recently, reweighting by another trial model utilizing self-learning techniques was proposed to reduce the autocorrelation time efficiently. [10]

In this work, we focus on another kind of widely used reweighting technique, Wang-Landau (WL) algorithm [11, 12], and apply it to the determinant quantum Monte Carlo (DQMC) [13, 14]. The WL algorithm

provides a general and powerful strategy to obtain the flat-histogram simulation in the space of energy or some other given quantities for classical problems. For quantum models, the WL method has been successfully incorporated to the stochastic series expansion [15], diagrammatic Monte Carlo [16] and Renyi entropy calculations [17]. In this work, we take another approach by directly generalizing the energy space in the classical case to the “configuration weight space” ( $w$ -space) in the quantum case. Based on the WL method, the density of states  $\rho(w)$  in the  $w$ -space can be obtained with high accuracy, which results in a flat-histogram sampling in the  $w$ -space by choosing  $\rho^{-1}(w)$  as the MC sampling weight. Comparing to the usual importance (*e.g.* Metropolis or heat-bath) sampling, the autocorrelation time  $\tau_a$  can be greatly reduced, although at the price of sacrificing some sampling efficiency. The advantage of such a WL assisted DQMC (WL-DQMC) algorithm is that some problems with super long autocorrelation times previously can now be reliably simulated by multi short walkers with high accuracy as a result of the reduced  $\tau_a$ .

As an application of our algorithm, we choose the two-dimensional half-filled Holstein model,

$$H = -t \sum_{\langle ij \rangle \sigma} (c_{i\sigma}^\dagger c_{j\sigma} + \text{h.c.}) + \Omega \sum_i b_i^\dagger b_i + \eta \sum_i (n_i - 1)(b_i^\dagger + b_i), \quad (1)$$

where  $c_{i\sigma}$  and  $b_i$  are electron and phonon field operators,  $t$  is the hopping taken as 1 below,  $\Omega$  is the phonon frequency, and  $\eta$  is the electron-phonon coupling. As usual, we define the dimensionless electron-phonon constant  $\lambda = 2\eta^2/\Omega W$  where  $W$  is the band width of the free fermion model. The Holstein model can be simulated using the standard DQMC method by sampling the displacement field directly [18], which is free of sign problem but notorious for its super long autocorrelation time. [19] However, recently, it received more and more attention for its underlying interesting properties such as charge density wave (CDW), superconductivity, and quantum critical point. [20–23] Recently, there have also been some other QMC approaches to the Holstein or

\* dawang@nju.edu.cn

† qhwang@nju.edu.cn

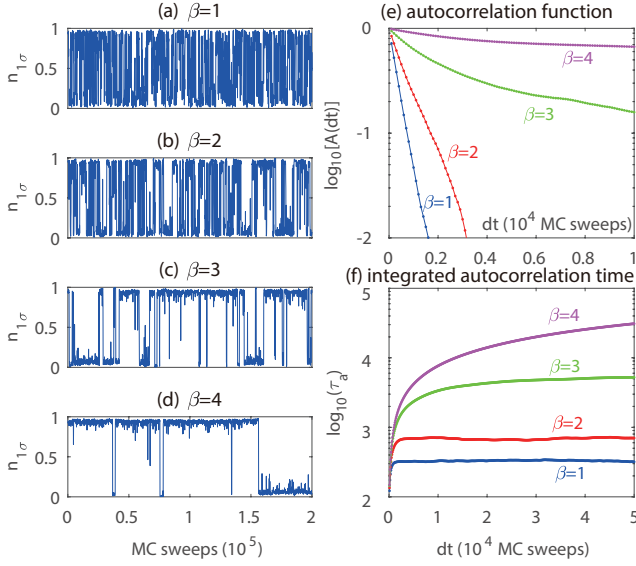


FIG. 1. DQMC simulations of the Holstein model on the square lattice with  $L = 4$ ,  $\Omega = 1$  and  $\lambda = 0.5$ . The discrete time slice  $\Delta\tau = 0.1$  in all simulations in this work. (a) to (d) plot the histogram of the particle number  $n_{1\sigma}$ . The normalized autocorrelation function  $A(dt)$  of  $n_{1\sigma}$  is shown in (e), from which we extract the integrated autocorrelation times  $\tau_a$  as defined in Eq. 3 plotted in (f).

Holstein-Hubbard model, including the interaction expanded continuous time QMC [24–27], machine learning assisted DQMC [28, 29], hybrid QMC [30], Langevin QMC [31, 32]. In this work, we are intended to incorporate the WL algorithm to the standard DQMC and test it on the Holstein model.

We first show the ergodicity problem in the usual DQMC simulations on the square lattice. In Fig. 1(a)-(d), we plot the histogram of the particle number at one site  $n_{1\sigma}$  from  $\beta = 1$  to 4, respectively. Clearly, as  $\beta$  increases, the change (from  $\sim 1$  to  $\sim 0$  or vice versa) of  $n_{1\sigma}$  becomes more and more difficult, indicating the ergodicity problem. It can be quantitatively characterized by the normalized autocorrelation function

$$A(dt) = \frac{\langle n_{1\sigma}(t)n_{1\sigma}(t+dt) \rangle}{\langle n_{1\sigma}^2 \rangle}. \quad (2)$$

In Fig. 1(e),  $A(dt)$  is plotted versus  $dt$ , in which the exponential dependence of  $A(dt) \sim e^{-dt/\tau_a}$  is obtained. In this work, we follow the standard definition of the integrated autocorrelation time  $\tau_a(dt)$  [3], *i.e.*

$$\tau_a(dt) = \frac{1}{2} + \sum_{k=1}^{dt} A(k), \quad (3)$$

which grows up as  $dt$  increases and finally saturates to a practical estimation of the autocorrelation time, as shown in Fig. 1(f). Roughly speaking,  $\tau_a$  exponentially depends on  $\beta$  reflecting the effect of the Boltzmann weight factor

$e^{-\beta\Delta}$  for tunneling across a barrier  $\Delta$ . In these simulations, the global update by shifting a distance for each site with all time slices [9, 18] has already been used, but the autocorrelation time is still too long. For the case of  $\beta = 4$ ,  $\tau_a$  is at the order of  $10^5$  MC sweeps, which demands the measurements more than  $\sim 10^7$  MC sweeps (much longer than  $\tau_a$ ) for each walker in a serious MC simulation. Clearly, as  $\beta$  increases, the MC simulations become more and more time consuming. Therefore, how to reduce the autocorrelation time is a very important problem.

In the DQMC method, by keeping physical boson fields or introducing auxiliary fields, the fermionic degrees of freedom can be integrated out, giving rise to the bosonic partition function  $Z = \sum_c w(c)$  where  $w(c)$  is in general can be negative or complex. Right now, we suppose real  $w(c) > 0$  for simplicity. The generalization to the cases with sign problem is straightforward. Defining the density of states (DOS)  $\rho(w) = \sum_c \delta[w - w(c)]$ , the partition function can be rewritten as  $Z = \sum_w w\rho(w)$ . The DOS  $\rho(w)$  can be obtained using the standard WL algorithm. After  $\rho(w)$  is obtained with high accuracy, we obtain a flat-histogram sampling in the  $w$ -space by choosing  $\rho^{-1}[w(c)]$  as the Markov chain weight. Then, a physical quantity  $\langle A \rangle$  becomes

$$\begin{aligned} \langle A \rangle &= \frac{\sum_c A(c)w(c)}{\sum_c w(c)} = \frac{\sum_c A(c)w(c)\rho[w(c)]\rho^{-1}[w(c)]}{\sum_c w(c)\rho[w(c)]\rho^{-1}[w(c)]} \\ &\rightarrow \frac{\sum'_i A_i w_i \rho_i}{\sum'_i w_i \rho_i} = \frac{\langle Aw\rho \rangle}{\langle w\rho \rangle} \end{aligned} \quad (4)$$

where  $\sum'$  means summation on the Markov chain. In practice, we record  $\ln \rho$  and  $\ln w$  (corresponding to the action) in computers to avoid numerical overflow. The algorithm is now described as follows:

Step-1. One performs a short DQMC run at first to get the range of  $\ln w$  in the usual DQMC sampling.

Step-2. By choosing a larger WL-window  $\ln w \in [W_{\min}, W_{\max}]$  covering the obtained  $\ln w$ -range in step-1 to perform the standard WL algorithm: we accept a new configuration  $c'$  with the probability  $r = \min[\rho[w(c)]/\rho[w(c')], 1]$ . Since the weight ratio is already obtained in DQMC, this step causes no additional computational effort.

Step-3. For each sampled configuration with weight  $w$ , we update  $\ln \rho(\ln w) = \ln \rho(\ln w) + \ln f$ , where  $f$  is initially chosen as  $e$ . After the histogram is “almost” flat, we set  $f \rightarrow \sqrt{f}$  and repeat steps-2 and -3 until  $f$  is sufficiently close to 1, say  $f < 1 + 10^{-6}$ . Then, the desired DOS  $\ln \rho(\ln w)$  is obtained.

Step-4. We stop updating  $\ln \rho$  and begin to do measurements using  $\rho^{-1}$  as the sampling weight. Notice that according to Eq. 4, we need accumulate  $A_i w_i \rho_i$  and  $w_i \rho_i$  simultaneously to get the physical quantity  $\langle A \rangle = \langle Aw\rho \rangle / \langle w\rho \rangle$ .

As an example to clarify the algorithm, let us focus on the square lattice with  $L = 4$ ,  $\lambda = 0.5$  and  $\beta = 4$ .

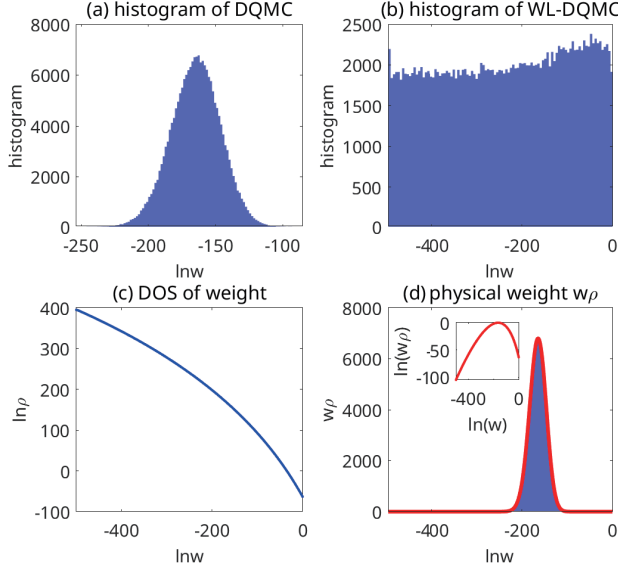


FIG. 2. Histograms of the configuration weight  $\ln w$  are plotted in (a) for DQMC and (b) for WL-DQMC. (c) DOS  $\ln \rho$  obtained by WL-DQMC versus  $\ln w$ . (d) By multiplying  $w$  and  $\rho$ , we obtain the physical weight which is the same as in (a) shown as the shaded area.

With this parameter setup, we will show that the autocorrelation time  $\tau_a$  is neither too short (standard DQMC applies) nor too long (WL-DQMC also fails) but in a reasonable range: of order  $10^5$  MC sweeps in DQMC even with the global update. At first, we perform a usual DQMC run. The histogram is plotted in Fig. 2(a), which tells us the physically important region is almost  $\ln w \in [-250, -100]$ . Then by choosing a larger WL-window  $[-500, -50]$ , we perform the WL sampling to obtain the DOS  $\ln \rho$  versus  $\ln w$  as shown in Fig. 2(c). Interestingly, the DOS is not bounded in the WL-window and continues to grow up as  $\ln w$  decreases. This behavior is not changed for moderately larger WL-windows. In fact, for the Holstein model, there should be no lower bound of  $\ln w$  since the phonon displacement can be arbitrarily large. With the obtained  $\rho^{-1}$  as the sampling weight, we do obtain the desired flat histogram as shown in Fig. 2(b). By multiplying  $w$  and  $\rho$ , we can reconstruct the “physical weight” which is in good agreement with the usual importance sampling, as shown in Fig. 2(d).

At first glance, the WL-DQMC enlarges the sampling region in the  $w$ -space and thus performs many “useless” measurements, which inevitably decreases the sampling efficiency. However, the lower-weight configurations (corresponding to higher-energy configurations classically) can reduce the tunneling barrier between different local minimums in the phase space. As depicted in the schematic diagram Fig. 3, the tunneling probability  $e^{-\beta\Delta}$  in the usual DQMC can be greatly enlarged to  $e^{-\beta\Delta'}$  by the WL-DQMC sampling. Therefore, the autocorrelation

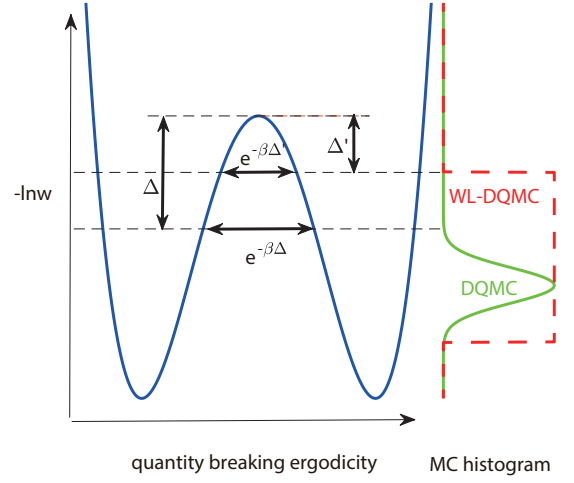


FIG. 3. A schematic plot showing how WL-DQMC increases the tunneling probability and thus reduces the autocorrelation time for exponential barrier problems.  $-\ln w$  (action) takes the role of free energy in the classical case. The MC histogram of WL-DQMC is plotted with solid green (dashed red) curves on the right.

time is greatly reduced. In fact, it can be expected that if the WL-window is large enough, the tunneling gap  $\Delta'$  vanishes and the transition probability only depends on  $\beta$  polynomially. Comparing with the usual reweighting technique with a higher temperature[6] or another group of parameters, the WL-DQMC has its advantage that the most important configurations are fully kept, while the sampling efficiency (roughly the area of DQMC over WL-DQMC in the histogram in Fig. 3) only decreases inversely proportional to the WL-window size.

The above picture is fully consistent with our numerical results. As shown in Fig. 4(a)-(d), by decreasing the lower bound of the WL-window  $W_{\min}$  while keeping  $W_{\max} = -50$ , we indeed observed that the change of  $n_{1\sigma}$  occurs more and more frequently. Correspondingly, the integrated autocorrelation time  $\tau_a$  drops significantly up to two orders (for  $W_{\min} = -2000$  to  $\sim 10^3$  MC sweeps) relative to the usual DQMC, as shown in Fig. 4(e). Of course, the price of the decrease of  $\tau_a$  is the drop of  $\langle w\rho \rangle$ , as shown in Fig. 4(f). However, as long as  $\langle w\rho \rangle$  is not too small, say  $\langle w\rho \rangle > 10^{-3}$ , we can still get reliable data by the means of statistical (multi-walker) average. The total efficiency of the WL-DQMC algorithm is determined by these two factors:  $\tau_a$  and  $\langle w\rho \rangle$ . But as  $\tau_a$  can be reduced exponentially while  $\langle w\rho \rangle$  only drops algebraically, their combination is still anticipated to be efficient.

As another application, we turn to the half-filled honeycomb lattice which was extensively studied very recently. [22, 23] We choose one typical group of parameters:  $L = 4$ ,  $\lambda = 2/3$ , and  $\beta = 8$  for comparing with Ref. 23. As shown in Fig. 5(a), the particle number of one site  $n_{1\sigma}$  is pinned at  $\sim 1$  within  $10^7$  MC sweeps. Similarly, the CDW form factor  $F_{\text{CDW}} =$

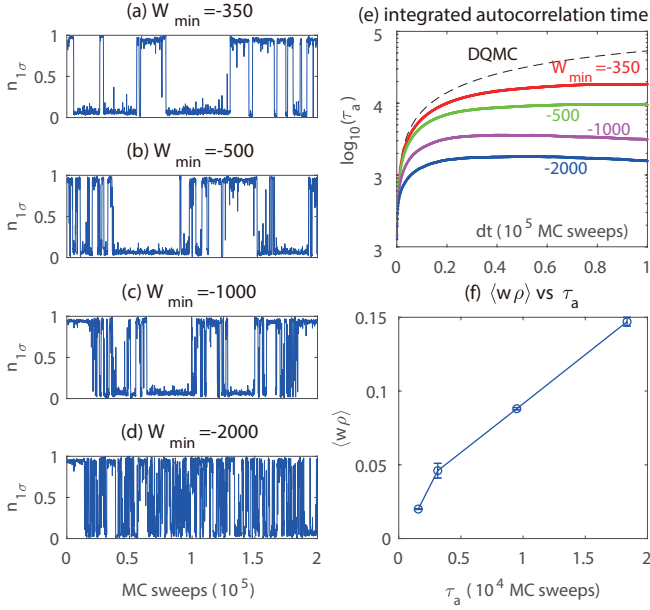


FIG. 4. Reducing autocorrelation time using the WL-DQMC with different WL-windows  $[W_{\min}, -50]$ . (a)-(d) plot the histograms of  $n_{1\sigma}$  with different  $W_{\min}$ . The integrated autocorrelation time  $\tau_a$  is shown in (e) together with the DQMC result for comparison. In (f), we plot the average  $\langle w\rho \rangle$  versus  $\tau_a$ , which clearly shows the decrease of  $\tau_a$  simultaneously accompanies the decrease of  $\langle w\rho \rangle$ .

$\frac{1}{2L^2} \sum_i (-1)^{i+j} \langle n_i n_j \rangle$  is also pinned at values from 24 to 26 as shown in Fig. 5(b). This is a clear feature of the ergodicity breaking and indicates that the autocorrelation time  $\tau_a$  should be at least larger than  $10^7$  MC sweeps, causing great difficulty in MC simulations. As an approximation, the researchers performed simulations with multi short runs to do average. [23] However, the approximation error is difficult to estimate. By trying different WL-windows, we found  $[W_{\min}, W_{\max}] = [-2000, -500]$  can reduce the autocorrelation times of  $n_{1\sigma}$  and  $F_{CDW}$  to  $\sim 1.5 \times 10^5$  and  $\sim 3 \times 10^4$  MC sweeps (not shown), respectively. This now enabled us to perform more reliable MC simulations. In Fig. 5(e), we plot the results of  $\langle F_{CDW} \rangle$  versus MC sweeps in ten walkers using WL-DQMC and DQMC, respectively. As anticipated by the shorter  $\tau_a$ , we did observe that the ten WL-DQMC walkers almost converge at order of  $10^6$  MC sweeps. Their difference comes mainly from statistical error. In contrast, the ten DQMC walkers give quite inconsistent results up to  $10^7$  MC sweeps as a result of the super long autocorrelation time. It should be mentioned that if we take the “false convergence” [six of ten walkers, indicated by the arrow in Fig. 5(e)] as an approximation, the error is found to be small. Therefore, we conclude that in this specific model the approximated DQMC simulations with running time much shorter than autocorrelation time [23]

are still qualitatively correct.

In summary, we implemented the WL algorithm in the

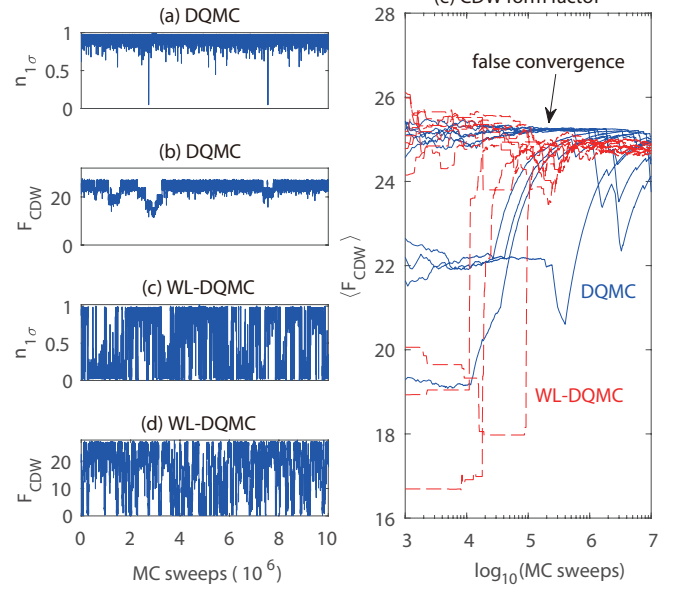


FIG. 5. The WL-DQMC simulations on the honeycomb lattice with  $L = 4$ ,  $\lambda = 2/3$ , and  $\beta = 8$ . (a) and (b) plot the histograms of  $n_{1\sigma}$  and  $F_{CDW}$  obtained in the usual DQMC simulations with the global update included. Similarly, (c) and (d) are obtained in our WL-DQMC algorithm with WL-window  $[W_{\min}, W_{\max}] = [-2000, -500]$ . In (e), the MC averaged CDW form factor  $F_{CDW}$  are plotted versus the MC sweeps in ten walkers for DQMC (solid blue lines) and WL-DQMC (dashed red lines), respectively.

DQMC framework to realize the flat-histogram sampling in the configuration weight space. The advantage of such a WL-DQMC method is to greatly reduce the autocorrelation time for exponential barrier problems. Its feasibility is checked in the Holstein model on both square and honeycomb lattices. With the WL-DQMC algorithm, some problems with long autocorrelation times previously can now be simulated possibly. Of course, our present work is a direct application of the WL algorithm in the DQMC by replacing the energy space to configuration weight space. Clearly, such an idea can also be applied to other quantum Monte Carlo algorithms as long as the flat-histogram is suitably chosen. As a future direction, we can achieve the flat-histogram in the space of some other physical quantities, *e.g.* order parameters which may be more efficient to recover the ergodicity.

D.W. thanks Zhichao Zhou, Yuxi Zhang, Tianxing Ma, Chuang Chen and Zi-Yang Meng for helpful discussions. This work is supported by the NSFC under Grant Nos. 11874205 and 11574134. The numerical simulations were performed in High-Performance Computing Center of Collaborative Innovation Center of Advanced Microstructures, Nanjing University.

- 
- [1] M. Troyer and U.-J. Wiese, *Phys. Rev. Lett.* **94**, 170201 (2005).
- [2] This is not in contradict with the ergodicity theorem which only states that the time average when  $T \rightarrow \infty$  is equal to the ensemble (very huge phase space) average but not (only a few) walker average.
- [3] W. Janke, in *Computational many-particle physics* (Springer, 2008) pp. 79–140.
- [4] R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **58**, 86 (1987).
- [5] U. Wolff, *Phys. Rev. Lett.* **62**, 361 (1989).
- [6] R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).
- [7] B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).
- [8] K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [9] R. T. Scalettar, R. M. Noack, and R. R. P. Singh, *Phys. Rev. B* **44**, 10502 (1991).
- [10] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, *Phys. Rev. B* **95**, 041101 (2017).
- [11] F. Wang and D. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- [12] F. Wang and D. P. Landau, *Phys. Rev. E* **64**, 056101 (2001).
- [13] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, *Phys. Rev. D* **24**, 2278 (1981).
- [14] F. F. Assaad and H. G. Evertz, in *Computational Many-Particle Physics* (Springer, 2008) pp. 277–356.
- [15] M. Troyer, S. Wessel, and F. Alet, *Phys. Rev. Lett.* **90**, 120201 (2003).
- [16] N. G. Diamantis and E. Manousakis, *Phys. Rev. E* **88**, 043302 (2013).
- [17] S. Inglis and R. G. Melko, *Phys. Rev. E* **87**, 013306 (2013).
- [18] S. Johnston, E. A. Nowadnick, Y. F. Kung, B. Moritz, R. T. Scalettar, and T. P. Devereaux, *Phys. Rev. B* **87**, 235133 (2013).
- [19] M. Hohenadler and T. C. Lang, in *Computational Many-Particle Physics* (Springer, 2008) pp. 357–366.
- [20] N. C. Costa, T. Blommel, W.-T. Chiu, G. G. Batrouni, and R. T. Scalettar, *Phys. Rev. Lett.* **120**, 187003 (2018).
- [21] S. Karakuzu, K. Seki, and S. Sorella, *Phys. Rev. B* **98**, 201108 (2018).
- [22] C. Chen, X. Y. Xu, Z. Y. Meng, and M. Hohenadler, *Phys. Rev. Lett.* **122**, 077601 (2019).
- [23] Y.-X. Zhang, W.-T. Chiu, N. Costa, G. Batrouni, and R. Scalettar, *Phys. Rev. Lett.* **122**, 077602 (2019).
- [24] F. F. Assaad and T. C. Lang, *Phys. Rev. B* **76**, 035116 (2007).
- [25] M. Weber, F. F. Assaad, and M. Hohenadler, *Phys. Rev. Lett.* **119**, 097401 (2017).
- [26] M. Weber and M. Hohenadler, *Phys. Rev. B* **98**, 085405 (2018).
- [27] M. Hohenadler and G. G. Batrouni, *Phys. Rev. B* **100**, 165114 (2019).
- [28] C. Chen, X. Y. Xu, J. Liu, G. Batrouni, R. Scalettar, and Z. Y. Meng, *Phys. Rev. B* **98**, 041102 (2018).
- [29] S. Li, P. M. Dee, E. Khatami, and S. Johnston, *Phys. Rev. B* **100**, 020302 (2019).
- [30] S. Beyl, F. Goth, and F. F. Assaad, *Phys. Rev. B* **97**, 085144 (2018).
- [31] G. G. Batrouni and R. T. Scalettar, *Phys. Rev. B* **99**, 035114 (2019).
- [32] B. Cohen-Stead, K. Barros, Z. Meng, C. Chen, R. T. Scalettar, and G. G. Batrouni, *Phys. Rev. B* **102**, 161108 (2020).