Accuracy and convergence of the Wang-Landau sampling algorithm

Alexander N. Morozov and Sheng Hsien Lin
Institute of Atomic and Molecular Sciences, Academia Sinica, P.O. Box 23-166, Taipei, Taiwan, Republic of China
(Received 5 March 2007; revised manuscript received 27 June 2007; published 1 August 2007)

We present estimations of the accuracy and convergence of the Wang-Landau algorithm. Both accuracy and the related length of the Monte Carlo run depend on the modification parameter f and the density of states. The analytical solution obtained for the two-level system was checked numerically on the two-dimensional Ising model. Although the two-level system is a very simple one, it appears that the proposed solution describes the generic features of the Wang-Landau algorithm. The estimations should prove useful in Monte Carlo calculations of protein folding, first-order transitions, and other systems with a rough energy landscape.

DOI: 10.1103/PhysRevE.76.026701 PACS number(s): 05.10.Ln, 05.50.+q

I. INTRODUCTION

Wang and Landau (WL) proposed an elegant and efficient Monte Carlo (MC) method for a direct estimation the density of states $g(E)$ using a random walk in energy space [1,2]. The WL algorithm offers quite a general way of sampling systems with a complicated energy landscape while traditional MC techniques fail due to trapping in local energy minima. The resultant $g(E)$ distribution can be used to calculate canonical averages of thermodynamic quantities at any temperature. Further studies and generalizations of the WL algorithm have been carried out by several authors [3–7]. A number of applications of the WL algorithm in protein folding, fluid simulations, systems with first-order phase transitions, and other systems with a rough energy landscape are thriving (see the review in Ref. [8]). One is interested in how to estimate the accuracy and how to control the convergence of WL algorithm. The proof of convergence was proposed by Zhou and Bhatt [9]. They have found the statistical error to scale as $\sqrt{\ln f}$ with the modification parameter f. On the other hand, the statistical error must depend on the tunneling time which is defined as the average number of MC steps to move between the lowest and highest energy levels of the system. Using a perfect flat-histogram scheme, Dayal *et al.* [10] showed that the system size and the specific properties of the $g(E)$ distribution determine the scaling behavior of the tunneling time. In a further advance Shell *et al.* [11] showed that the tunneling time scale is defined uniquely by the entropy range of the energy window sampled. The proposed analytical consideration relates accuracy and convergence with the simulation parameters of the WL algorithm and the properties of the $g(E)$ distribution. It should be noted that convergence can be controlled by the "flatness" of the histogram, as was proposed in the original works [1,2]. However, the flatness criterion is not strict and depends on the researcher's choice.

The WL stochastic process does not satisfy detailed balance and cannot be approximated by a Marcovian chain, making it a complicated task to reach our goals if one considers the general case. In this paper we present approximate solutions for the simple two-level system. The results become accurate in the limit $\ln f=\varepsilon \ll 1$ and can be used as the approximation for systems with multilevel energy spectra. Although the two-level system is a very simple one, it appears that the mechanism of WL convergence is essentially the same as in complex systems. The results were checked numerically on the two-dimensional (2D) Ising models. Since the WL method offers substantial advantages over traditional MC techniques [1,2], the proposed estimations could be of considerable value. The possible strategies for optimal performance in parallelized WL algorithms are discussed in the final part of the paper.

II. TWO-LEVEL SYSTEM

We shall first describe briefly the WL algorithm for systems with discrete multilevel energy spectra $\{E_1,\ldots,E_m,\ldots,E_M\}$. In this case $g(E_m)$ designates the number of states for the energy level E_m. The WL algorithm produces the stochastic process $\mathbf{x}_1\rightarrow\mathbf{x}_2\cdots\rightarrow\mathbf{x}_{l-1}\rightarrow\mathbf{x}_l$ in the conformational space $\mathbf{x}$ with a transition probability which is proportional to the reciprocal of $g(E)$ value:

$$w(\mathbf{x}\rightarrow\mathbf{x}')=\min\left[\frac{g(E(\mathbf{x}))}{g(E(\mathbf{x}'))},1\right]. \quad (1)$$

After each MC step the density of states $g(E)$ and the histogram $H(E)$ are updated:

$$g(E(\mathbf{x}_l))\rightarrow g(E(\mathbf{x}_l))f_i,\quad H_i(E(\mathbf{x}_l))\rightarrow H_i(E(\mathbf{x}_l))+1, \quad (2)$$

where the subscript i denotes the ith iteration of the WL algorithm. At the beginning of the first iteration the initial modification factor f_1 and initial approximation $g_1(E)$ for the density of states have to be chosen. At the beginning of the ith iteration the modification factor is reduced; i.e., $f_i=(f_{i-1})^{1/a}$ and $H_i(E)=0$ have to be reset for all energy levels. Finally, if $i\rightarrow\infty$ and the WL stochastic chain is long enough, a flat histogram can be obtained, $H(E_m)\cong$const, $m=1,\ldots,M$, while $g_i(E)$ converges to within a multiplicative constant to the exact distribution of the density of states. In the original works [1,2] of Wang and Landau, $f_1=e=2.7182\ldots$, $g_1(E)=1$, and $a=2$ were chosen for the 2D Ising model.

Let us consider a simple system with just two levels of energy $\{E_1,E_2\}$. For a randomly chosen conformation the probability to occupy one of the levels reads

 ©2007 The American Physical Society

$$p = \frac{g(E_1)}{g(E_1)+g(E_2)}, \quad q = \frac{g(E_2)}{g(E_1)+g(E_2)}, \tag{3}$$

where one can limit consideration to the case $g(E_2) \geqslant g(E_1)$. Further, we designate the number of states due to the WL algorithm as g_1 and g_2 to differ from the exact values of $g(E_1)$ and $g(E_2)$. We define the deviation Δ from the exact values using the formula

$$\frac{g_2}{g_1} = \frac{g(E_2)}{g(E_1)} \exp(\Delta) = \frac{q}{p} \exp(\Delta). \tag{4}$$

In the WL algorithm the probabilities of transitions between the E_1 and E_2 levels are given by

$$w^{2\to 1} = p \min\left[\frac{g_2}{g_1}, 1\right], \quad w^{1\to 2} = q \min\left[\frac{g_1}{g_2}, 1\right]. \tag{5}$$

The probabilities of transitions to occur after m steps read

$$w_m^{2\to 1} = p s_m^{2\to 1} \prod_{k=1}^{m-1} (1 - p s_k^{2\to 1}),$$

$$w_m^{1\to 2} = q s_m^{1\to 2} \prod_{k=1}^{m-1} (1 - q s_k^{1\to 2}),$$

$$s_k^{2\to 1} = \min\left[\frac{g_2}{g_1} f_i^{k-1}, 1\right], \quad s_k^{1\to 2} = \min\left[\frac{g_1}{g_2} f_i^{k-1}, 1\right], \tag{6}$$

where g_1 and g_2 are the number of states when $m=1$ and the product $\prod_{k=1}^{m-1}(\cdots)$ is equal to unity if $m=1$. Let us define an event as the consecutive $2\to 1\to 2$ transition and consider a chain of these events $\{1,\ldots,n-1,n,\ldots\}$. Using Eqs. (2) and (4) one can find

$$\begin{aligned}\Delta_n &= \Delta_{n-1} + (m_n^{2\to 1} - m_n^{1\to 2}) \ln f_i \\ &= \Delta_{n-1} + (\langle m_n^{2\to 1}\rangle - \langle m_n^{1\to 2}\rangle) \ln f_i + (\delta m_n^{2\to 1} - \delta m_n^{1\to 2}) \ln f_i,\end{aligned} \tag{7}$$

where $m_n^{2\to 1}$ and $m_n^{1\to 2}$ are the stochastic numbers which were realized at the nth event due to the probability distributions (6), and $\delta m_n^{2\to 1}$ and $\delta m_n^{1\to 2}$ are the deviations from their average values. In Eq. (7), Δ_0 describes the error due to the previous iteration of the WL algorithm. Let us first consider the case when the value of Δ_{n-1} satisfies the inequality

$$\Delta_{n-1} + \ln\left(\frac{q}{p}\right) \geqslant 0. \tag{8}$$

In this case $s_m^{2\to 1} = 1$ and one can find the average value of $m_n^{2\to 1}$ from the geometric distribution:

$$\langle m_n^{2\to 1}\rangle = \sum_{m=1}^{\infty} m w_m^{2\to 1} = p \sum_{m=1}^{\infty} m (1-p)^{m-1} = \frac{1}{p}. \tag{9}$$

For $1\to 2$ transitions the WL factor has to be taken into account:

$$\begin{aligned} m \leqslant m_{\max} = 1 + m_n^{2\to 1} + \frac{\ln(q/p) + \Delta_{n-1}}{\ln f_i} &\Rightarrow s_m^{1\to 2} \\ = \frac{p f_i^{m-1}}{q} \exp(-\Delta_{n-1} - m_n^{2\to 1} \ln f_i), & \\ m > m_{\max} \Rightarrow s_m^{1\to 2} = 1. & \end{aligned} \tag{10}$$

The transition probability $w_m^{1\to 2}, m \leqslant m_{\max}$ can be represented in the form

$$w_m^{1\to 2} = q s_m^{1\to 2} \prod_{k=1}^{m-1} (1 - q s_k^{1\to 2}) = \xi (1-\xi)^{m-1} f_i^{m-1} \theta_m(\xi, f_i), \tag{11}$$

where $\xi = p \exp(-\Delta_{n-1} - m_n^{2\to 1} \ln f_i)$ and $\theta_m(\xi, f_i) = \prod_{k=1}^{m-1}\left[1 - \frac{\xi}{1-\xi}(f_i^{k-1} - 1)\right]$. One can see that in the limit $\ln f_i = \varepsilon_i \to 0$ it follows that $m_{\max} \to \infty$, $f_i^{m-1}\theta_m(\xi, f_i) \to 1$ and consequently the transition probability is given by the geometric distribution $w_m^{1\to 2} = \xi_0 (1-\xi_0)^{m-1}$, $\xi_0 = p \exp(-\Delta_{n-1})$. The average of $m_n^{1\to 2}$ can be expanded over ε_i:

$$\langle m_n^{1\to 2}\rangle = \left(\sum_{m=1}^{\infty} m w_m^{1\to 2}\right) \Bigg/ \left(\sum_{m=1}^{\infty} w_m^{1\to 2}\right) = \frac{1}{\xi_0} + \sum_{k=1}^{\infty} a_k^{(n)} \varepsilon_i^k. \tag{12}$$

In Eq. (12) the factors $a_k^{(n)}$ depend on $m_n^{2\to 1}$ and Δ_{n-1}. The series $\Sigma_{k=1}^{\infty} a_k^{(n)} \varepsilon_i^k$ has the upper and lower bounds (see Appendix A). Therefore in the limit $\varepsilon_i \to 0$ the leading term of the ε_i expansion in Eq. (12) is given by $a_1^{(n)}\varepsilon$, where $a_1^{(n)}$ factor reads (see Appendix A)

$$a_1^{(n)} = \frac{1}{\xi_0}\left(m_n^{2\to 1} + \frac{7\xi_0 - 8}{2\xi_0}\right). \tag{13}$$

Substituting Eqs. (9) and (12) into Eq. (7) gives

$$\begin{aligned}\Delta_n = \Delta_{n-1} + \frac{1 - \exp(\Delta_{n-1})}{p}\varepsilon_i + (\delta m_n^{2\to 1} - \delta m_n^{1\to 2})\varepsilon_i \\ - \sum_{k=1}^{\infty} a_k^{(n)} \varepsilon_i^{k+1}.\end{aligned} \tag{14}$$

From Eq. (14) it follows in the limit $\varepsilon_i \to 0$ that the average of Δ_n over all possible $m_n^{2\to 1}, m_n^{1\to 2}$ pairs tends to zero. Moreover, the expected value of Δ_n to the kth power obeys to the next inequality (see Appendix B)

$$|\langle \Delta_n^k \rangle| < |\langle \Delta_{n-1}^k \rangle|, k = 1,2,\ldots . \tag{15}$$

Hence Δ_n is attracted to zero if $\varepsilon_i \to 0$ and without losing generality one can use the linear over Δ approximation of Eq. (14):

$$\Delta_n \cong \Delta_{n-1}\left(1 - \frac{\varepsilon_i}{p}\right) + (\delta m_n^{2\to 1} - \delta m_n^{1\to 2})\varepsilon_i - \sum_{k=1}^{\infty} a_k^{(n)} \varepsilon_i^{k+1}. \tag{16}$$

As long as the condition $\Delta_{n+l} + \ln(q/p) \geqslant 0$ is valid Eq. (16) can be iterated l times. From Eqs. (15) and (16) it follows

that l tends to infinity in the limit $\varepsilon_i \to 0$. The initial deviation Δ_0 can be chosen as Δ_{n-1} and after n iterations of Eq. (16) one can obtain

$$\Delta_n = \Delta_0\left(1-\frac{\varepsilon_i}{p}\right)^n + \varepsilon_i\sum_{l=1}^{n}(\delta m_l^{2\to 1} - \delta m_l^{1\to 2})\left(1-\frac{\varepsilon_i}{p}\right)^{n-l} - \sum_{l=1}^{n}\left(1-\frac{\varepsilon_i}{p}\right)^{n-l}\sum_{k=1}^{\infty} a_k^{(l)}\varepsilon_i^{k+1}. \quad (17)$$

For $n \gg 1$ the initial deviation Δ_0 vanishes and the error of the WL algorithm is due to the last two stochastic terms in Eq. (17). The expected value of $\Delta_n, n \gg 1$ can be estimated by averaging over all possible WL trajectories:

$$\langle\Delta_\infty\rangle = \sum_{k=1}^{\infty}\varepsilon_i^{k+1}\sum_{l=1}^{\infty}\langle -a_k^{(l)}\rangle\left(1-\frac{\varepsilon_i}{p}\right)^{n-l} \approx \varepsilon_i^2\sum_{l=1}^{\infty}\langle -a_1^{(l)}\rangle\left(1-\frac{\varepsilon_i}{p}\right)^{n-l} \leqslant \varepsilon_i \max\langle -a_1^{(l)}\rangle p, \quad (18)$$

where $\max\langle -a_1^{(l)}\rangle$ designates the maximum value of $\langle -a_1^{(l)}\rangle$. Using Eqs. (13) and (15) the inequality $\max\langle -a_1^{(l)}\rangle \leqslant 4/p^2 \max[\exp(2\Delta_0), 1]$ can be proven (see Appendix B).

Now we turn to the average deviation of Δ_∞ from $\langle\Delta_\infty\rangle$. The latter one can be estimated using the linear approximation over ε_i of Eq. (17):

$$\sigma(\Delta_\infty) \cong \sqrt{\left\langle\left[\varepsilon_i\sum_{l=1}^{\infty}(\delta m_l^{2\to 1} - \delta m_l^{1\to 2})(1-\varepsilon_i/p)^{n-l}\right]^2\right\rangle} \cong \varepsilon_i\sqrt{\frac{2\sigma^2(m^{2\to 1})}{1-(1-\varepsilon_i/p)^2}} \cong \sqrt{\varepsilon_i\frac{q}{p}} \cong \sqrt{\varepsilon_i\frac{g_2}{g_1}}, \quad (19)$$

where the relations $\langle\delta m_k^{2\to 1}\delta m_l^{2\to 1}\rangle = \langle\delta m_k^{1\to 2}\delta m_l^{1\to 2}\rangle = 0$ $k \neq l$, $\langle\delta m_k^{1\to 2}\delta m_l^{2\to 1}\rangle = 0$, $\langle(\delta m_k^{1\to 2})^2\rangle \approx \langle(\delta m_k^{2\to 1})^2\rangle = \sigma^2(m^{2\to 1}) = q/p^2$, and $(1-\varepsilon_i/p)^2 \approx 1-2\varepsilon_i/p$ were used (see Appendix C for details) to derive Eq. (19). Since $\sigma(\Delta_\infty) \gg \langle\Delta_\infty\rangle$, the average error of the WL algorithm in the limit $\varepsilon_i \to 0$ is given by Eq. (19). While the inequality (8) was used to obtain the estimations in Eqs. (18) and (19), a similar derivation can be done to show that these estimations provide the upper limit of the error for all the cases. The numerical calculations of the dispersion of the WL error for the two-level system with $q/p = e$ (see Fig. 1) are in a full agreement with Eq. (19).

The decay of the initial deviation in Eq. (17) provides the estimation of n that is sufficient for convergence of the WL algorithm:

$$\Delta_0\left(1-\frac{\varepsilon_i}{p}\right)^n = \varepsilon_i\frac{q}{p} \ll \sqrt{\varepsilon_i\frac{q}{p}} \Rightarrow n \cong \frac{p}{\varepsilon_i}\ln\Delta_0\frac{p/q}{\varepsilon_i}. \quad (20)$$

In the limit $\varepsilon_i \to 0$ the histogram is almost a flat distribution—that is, $n \cong H_i p$—whence one can find

$$H_i \geqslant \frac{1}{\varepsilon_i}\ln\Delta_0\frac{p/q}{\varepsilon_i} \cong \frac{1}{\varepsilon_i}\ln\Delta_0\frac{g_1/g_2}{\varepsilon_i}. \quad (21)$$

The numerical calculations of convergence for $\Delta_0 = 1$ and $q/p = e$ (see Fig. 1) are in a full agreement with Eq. (21). For

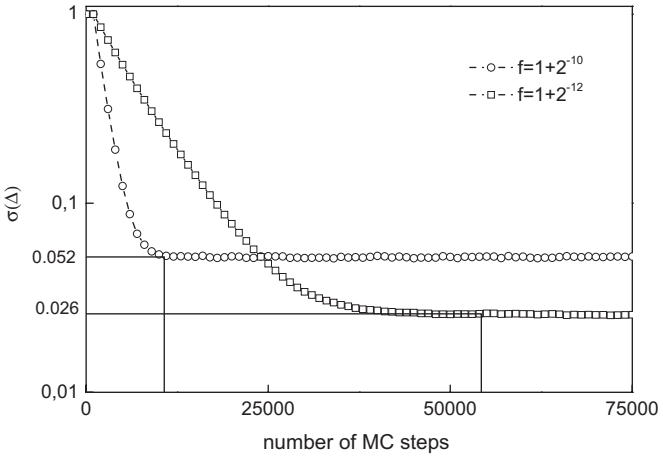

FIG. 1. Dashed lines with scatters: the dispersion of the error of the WL algorithm for the two-level system ($\Delta_0 = 1$, $q/p = e$), calculated by averaging over 10 000 stochastic trajectories, vs the number of MC steps. Solid lines: the related analytical calculations of accuracy and convergence using Eqs. (19) and (21). The related values of $\langle\Delta\rangle$ averaged over 10 000 trajectories are $\langle\Delta\rangle \cong 0.002$ for $f = 1+2^{-10}$ and $\langle\Delta\rangle \cong 0.001$ for $f = 1+2^{-12}$.

$i \gg 1$ the error of the initial guess of the density of states vanishes and Δ_0 is due to the stochastic terms in Eq. (17). Substituting Eq. (19) into Eq. (21) gives the criterion of convergence:

$$H_i \geqslant \frac{1}{\varepsilon_i}\ln\frac{\sqrt{\varepsilon_{i-1}g_1/g_2}}{\varepsilon_i} = \frac{1}{2\varepsilon_i}\ln a\frac{g_1/g_2}{\varepsilon_i}, \quad (22)$$

where the relation $f_i = (f_{i-1})^{1/a} \Rightarrow \varepsilon_i = \varepsilon_{i-1}/a$ was used to derive Eq. (22). The estimated error and convergence of the WL algorithm, respectively Eqs. (19) and (22), are accurate for the two-level system in the limit $\varepsilon_i \to 0$.

III. MULTILEVEL SYSTEMS

In the multilevel systems with the density of states given by the distribution $g(E_m)$, $m = 1, \ldots, M$ [for continuum systems the distribution $g(E_m)$ is the discretized approximation of the true density of states], the mechanism of the WL algorithm is more complex. A strict theoretical treatment of this case is complicated. However, the use of the two-level results in the multilevel case can be justified under a few assumptions. Similar to Eq. (3) one can introduce the local error of the WL algorithm as $\frac{g_m}{g_k} = \frac{g(E_m)}{g(E_k)}\exp(\Delta_{mk})$. Depending on the particular realization of a way to pick up the trial states, the distribution of the separation $|m-k|$ should be expected for the WL transitions. It is reasonable to assume that accuracy and convergence are determined by the transitions with the most probable separation $|m-k|$. From the previous consideration of the two-level system one can conclude that if transitions $m \to k$ or $k \to m$ are considered separately, the change of the error has no definite direction. The net decrease of the average error Δ_{mk} should be expected in the

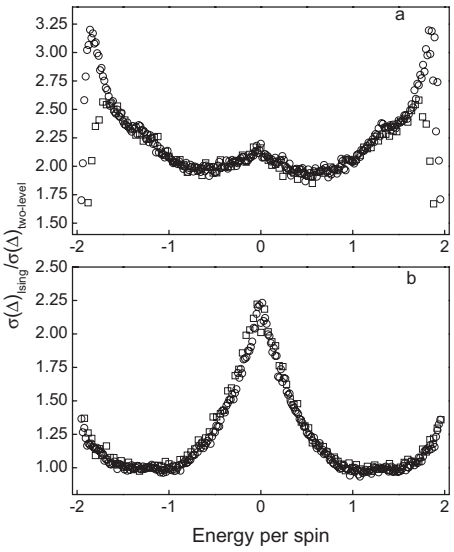

FIG. 2. The ratio $\sigma_m(\Delta)/\sqrt{\frac{g_m}{g_k}\varepsilon_i}$, $g_m \geqslant g_k$, for the 2D Ising model calculated with (a) $|m-k|=1$ and (b) $|m-k|=2$. $\sigma_m(\Delta)$ was averaged over 1600 WL trajectories for the two different sizes of the system: (□) -10×10 and (○) -16×16.

round-trip transitions—i.e., $m\to k\to m$. In this case Eqs. (19) and (22) for the two-level system with $g_1=g_k, g_2=g_m$ can be used as the approximations

$$\sigma_{i,m} \sim \sqrt{\gamma_m \varepsilon_i} \tag{23}$$

and

$$H_{i,m} \geqslant \frac{1}{2\varepsilon_i} \ln \frac{a}{\gamma_m \varepsilon_i}, \tag{24}$$

where $\gamma_m=g_m/g_k$, $g_m>g_k$ corresponds to the most probable value of $|m-k|$. Similar with the density of states the value of γ_m has to be found "on the fly." It should be noted that Eq. (23) with $\gamma_m=\max[g_m/g_{m-1}, g_m/g_{m+1}]$ provides the lower bound of the statistical error.

IV. RESULTS AND DISCUSSION

The estimations in Eqs. (23) and (24) were checked numerically on the 2D Ising model with the cell sizes 10×10 and 16×16. A flip of a single spin per MC step was used to update the configuration of the system. In this case the separation $|m-k|$ varies in the range $1\leqslant|m-k|\leqslant 2$. The values of $\sigma_{i,m}$ were calculated using the exact distribution of the density of states [12] by averaging over 1600 WL stochastic trajectories. The values $\sigma_{i,m}/\sqrt{\gamma_m\varepsilon_i}$ were calculated for $|m-k|=1,2$. For $|m-k|=1$ the result [see Fig. 2(a)] shows the dependence on the system size in the regions with low and high energies. These are the regions where the separation $|m-k|=2$ is the most probable. After recalculating with $|m-k|=2$ the ratio $\sigma_{i,m}/\sqrt{\gamma_m\varepsilon_i}$ does not depend on the system size [see Fig. 2(b)] and ranges from 1 to 2.2. It appears that the two-level approximation works better on the wings of the energy distribution. That is probably due to the sharper distribution of the separation $|m-k|$ in the regions with low and high energies. The estimation of convergence given by Eq. (24) is in good agreement with the numerical results too (see Fig. 3).

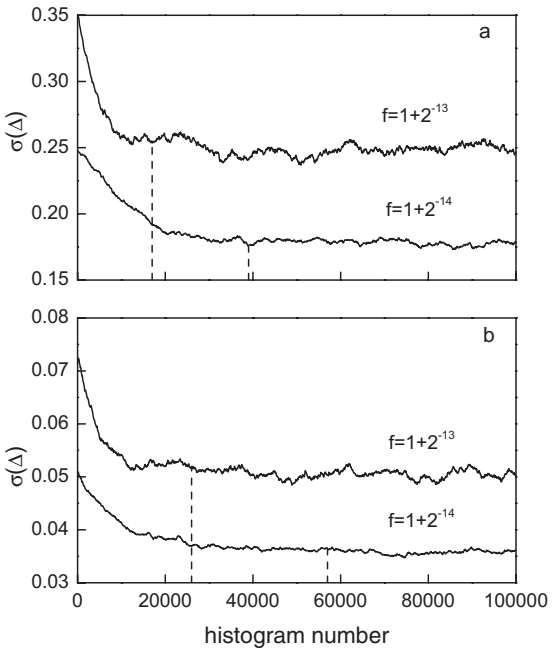

FIG. 3. Accuracy and convergence of WL algorithm: the 2D Ising model of finite size 16×16. Solid lines: the dispersion of the error, calculated by averaging over 1600 stochastic trajectories, vs the histogram number: (a) $E=-2$ and (b) $E=-1$ (E denotes energy per spin). Dashed lines: the related analytical calculations of convergence using Eq. (25): (a) $\gamma_m=256$, $a=2$ and (b) $\gamma_m\cong 25$, $a=2$.

The effect of the entropy gradient. The statistical error given by Eq. (24) is in accordance with the previous estimation $\sim\sqrt{\ln f}$ of Zhou and Bhatt [9]. The important new result is that the specific properties of $g(E)$ distribution have to be taken into account using in Eq. (23) the factor γ_m. It is useful to note that γ_m can be approximated by the gradient of the dimensionless entropy $S=\ln g$: that is,

$$\gamma_m = \frac{g_m}{g_k} = 1 + \frac{\Delta g_{km}}{g_k} \approx 1 + \frac{dS}{dE}\Delta E_{km}, \tag{25}$$

where all higher-order terms over ΔE_{km} were omitted in the expansion. Using Eqs. (23) and (25) it is instructive to explain the different limits of accuracy for the different energy levels [see Figs. 3(a) and 3(b)] as a result of the differences in the related entropy gradients. The lowest accuracy has to be expected for the ground state where the entropy gradient reaches the maximum value [Fig. 3(a)]. In a similar way the effect of system size on the statistical error can be understood as the result of a change of the entropy gradient with a change of the system size (see Fig. 4). In the 2D Ising model the entropy gradient for the ground state grows in direct proportion to the number of particles in the system. Accordingly, to maintain the given limit of accuracy one has to decrease the modification factor inversely as the number of particles with increasing system size. At the same time the statistical error for the energy level $E=0$ does not change substantially

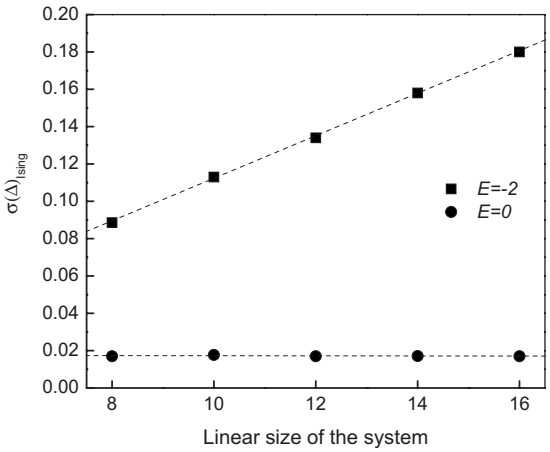

FIG. 4. The effect of the system size on the statistical error of the WL algorithm ($f=1+2^{-14}$). E denotes the energy per spin in the 2D Ising model.

with system size (see Fig. 4) in full agreement with the almost zero-entropy gradient at this level, independently of the system size.

Using Eqs. (9) and (12) the local tunneling time of $m \to k$ or $k \to m$ transition can be approximated as

$$\tau_{km} \approx \tau_{mk} = \frac{g_m + g_k}{g_k} \approx 2 + \frac{dS}{dE}\Delta E_{km} = 2 + \Delta S_{km}. \quad (26)$$

From Eq. (26) one can see that WL walks in the energy space are not true random walks. It takes more MC sweeps to make transitions in the regions with the higher-entropy gradient. Dayal *et al.* [10] found the deviation of the scaling behavior of the global tunneling time with the system size from that of an unbiased random walk in the energy space. To find the global tunneling time τ_{gl} between the bounds of the energy window sampled one should solve the diffusion equation for the probability of random walker [see Eq. (5) in Ref. [11]] but with the inhomogeneous diffusion coefficient which is defined by the local tunneling time. Equation (26) shows that the local tunneling time depends on the local entropy change in the transition. It sheds some light on the result obtained for the fluid systems by Shell *et al.* [11]: $\tau_{gl} \sim (\Delta S)^{2.2}$, where ΔS is the range of the entropy of the related energy window. Still it is a challenging problem for future study to explain the scaling laws in Refs. [10,11]. The solution of the nonhomogeneous diffusion equation is a possible way to solve the problem.

Effective parallelized WL algorithm. Shell *et al.* [11] pointed out that a key question of the parallelized WL algorithm is how to construct energy windows to load evenly the available processors and to reach simultaneously the given statistical error for all windows. On the basis of the scaling law $\tau_{gl} \sim (\Delta S)^{2.2}$ they have proposed to divide the complete energy range in the energy windows with an equivalent range of entropy. This approach can secure convergence at all windows approximately at the same time. However, according to our result, the homogeneity of the statistical error cannot be reached. One of the possible improvements is to combine the approach of Shell *et al.* [11] with that of Zhou and Bhatt [9]. According to Ref. [9] with K independent measurements of the histogram the statistical error is reduced to $\sigma_K \approx \sigma/\sqrt{K}$. Thus parallel measurements of the same window can compensate for an increase of the entropy gradient while simultaneous convergence for the different windows can be reached by proper division of the entropy range. Another possible approach is to control convergence and the change of the modification parameter for each window separately. In this case the nonhomogeneity of the statistical error can be compensated for by uneven division of the entropy range to reach faster convergence in regions with a higher-entropy gradient.

V. CONCLUSION

In conclusion, the WL algorithm provides quite a new and powerful approach in statistical physics. The estimations in Eqs. (23) and (24) allow control of the accuracy and convergence of the WL sampling algorithm. The results obtained are also useful for efficient implementations of the algorithm.

APPENDIX A

Because the updates of the density of states speed up the $1 \to 2$ transitions, the upper bound of $\langle m_n^{1\to 2}\rangle$ is given by the $w_m = \xi(1-\xi)^{m-1}$ distribution. The latter one corresponds to turning off the updates by setting $\theta_m f_i^{m-1} = 1$ in Eq. (11). Thus the $1 \leqslant \langle m_n^{1\to 2}\rangle \leqslant \xi^{-1}$ inequality is valid and consequently the $\Sigma_{k=1}^{\infty} a_k^{(n)} \varepsilon_i^k$ series has the upper and lower bounds and converges.

To find the $a_k^{(n)}$ factors in Eq. (12) it is convenient to use the recurrent relations for $\frac{d^n\theta_m}{d\varepsilon_i^n}$:

$$\frac{d\theta_m}{d\varepsilon_i} = \theta_m Y_m^{(1)}, \ldots, \frac{d^n\theta_m}{d\varepsilon_i^n} = \theta_m Y_m^{(n)}, \quad (A1)$$

$$Y_m^{(n)} = Y_m^{(1)} Y_m^{(n-1)} + \frac{dY_m^{(n-1)}}{d\varepsilon_i},$$

where $Y_m^{(1)} = \Sigma_{k=1}^{m-1} \frac{\xi(1-k)f_i^{k-1}}{1-\xi f_i^{k-1}}$. Let us find the factor $a_1^{(n)}$. Equation (A1) with $n=1$, $\varepsilon_i = 0$ gives

$$\left.\frac{d\theta_m}{d\varepsilon_i}\right|_{\varepsilon_i=0} = \left.\theta_m \sum_{k=1}^{m-1} \frac{\xi(1-k)f_i^{k-1}}{1-\xi f_i^{k-1}}\right|_{\varepsilon_i=0} = \frac{\xi}{1-\xi}\sum_{k=1}^{m-1}(1-k)$$

$$= \frac{\xi(m-1)^2}{2(\xi-1)}. \quad (A2)$$

Using Eq. (A2) one can obtain from Eq. (11) the relation for $w_m^{1\to 2}$:

$$w_m^{1\to 2} \cong \xi(1-\xi)^{m-1} f_i^{m-1}\left(1 + \frac{\xi(m-1)^2}{2(\xi-1)}\varepsilon_i\right). \quad (A3)$$

Substituting Eq. (A3) into Eq. (12) gives the expression for $\langle m_n^{1\to 2}\rangle$:

$$\langle m_n^{1\to 2}\rangle = \frac{\sum_{m=1}^{\infty} w_m^{1\to 2} + \frac{\xi(1-\xi)f_i}{[1-(1-\xi)f_i]^2} - \frac{\xi^2 f_i[1+4(1-\xi)f_i+(1-\xi)^2 f_i^2]}{2[1-(1-\xi)f_i]^4}\varepsilon_i}{\sum_{m=1}^{\infty} w_m^{1\to 2}}, \tag{A4}$$

where $\Sigma_{m=1}^{\infty} w_m^{1\to 2} = \xi[1-(1-\xi)f_i]^{-1} - \xi^2 f_i[1+(1-\xi)f_i]\{2[1-(1-\xi)f_i]^3\}^{-1}\varepsilon_i$. Using in Eq. (A4) the expansions $f_i \cong 1+\varepsilon_i$ and $\xi \cong \xi_0(1-m_n^{2\to 1}\varepsilon_i)$ one can obtain

$$\langle m_n^{1\to 2}\rangle \cong \frac{1}{\xi_0} + \frac{1}{\xi_0}\left(m_n^{2\to 1} + \frac{7\xi_0 - 8}{2\xi_0}\right)\varepsilon_i = \frac{1}{\xi_0} + a_1^{(n)}\varepsilon_i. \tag{A5}$$

The same approach can be used to find the higher-order factors $a_k^{(n)}$.

APPENDIX B

From Eq. (14) one can find to within ε_i accuracy that the value of Δ_n to the kth power reads

$$\Delta_n^k \cong \Delta_{n-1}^k + k\Delta_{n-1}^{k-1}\left[\frac{1-\exp(\Delta_{n-1})}{p}\varepsilon_i + (\delta m_n^{2\to 1} - \delta m_n^{1\to 2})\varepsilon_i\right]. \tag{B1}$$

The average over all possible $m_n^{2\to 1}, m_n^{1\to 2}$ pairs of Δ_n^k is given by

$$\langle \Delta_n^k\rangle = \langle \Delta_{n-1}^k\rangle + k\left\langle \Delta_{n-1}^{k-1}\frac{1-\exp(\Delta_{n-1})}{p}\right\rangle \varepsilon_i. \tag{B2}$$

One can check that the sign of the second term on the right-hand side of Eq. (B2) is always opposite to the sign of $\langle \Delta_{n-1}^k\rangle$. Therefore in the limit $\varepsilon_i \to 0$ the inequality in the expression (14) is valid. Let us prove the inequality

$$\max\langle -a_1^{(l)}\rangle \leqslant 4/p^2 \max[\exp(2\Delta_0), 1]. \tag{B3}$$

First, one can derive from Eq. (13) the relation

$$\langle -a_1^{(n)}\rangle < \frac{4}{p^2}\langle \exp(2\Delta_{n-1})\rangle. \tag{B4}$$

Then we shall consider the $\Delta_{n-1} > 0$ and $\Delta_{n-1} \leqslant 0$ cases. One can find from Eq. (16) that $\mathrm{sgn}(\Delta_{n-1}) = \mathrm{sgn}(\Delta_n)$ in the limit $\varepsilon_i \to 0$. From the inequality in expression (14) it follows that in the case $\Delta_{n-1} > 0$ the inequality $\langle \exp(2\Delta_n)\rangle < \langle \exp(2\Delta_{n-1})\rangle$ is valid. Hence the inequality $\langle \exp(2\Delta_n)\rangle < \exp(2\Delta_0)$ set the upper bound for $\langle \exp(2\Delta_n)\rangle$. In the case $\Delta_{n-1} \leqslant 0$ the inequality $\langle \exp(2\Delta_n)\rangle \leqslant 1$ is true. Finally the two cases are combined in expression (B3).

APPENDIX C

Let us introduce for convenience the binary variables $\alpha, \beta \in (0,1)$ and set the correspondence of $\alpha, \beta = 0$ to the $2 \to 1$ transition and $\alpha, \beta = 1$ to the $1 \to 2$ transition. The average value of $\langle \delta m_k^\alpha \delta m_l^\beta\rangle$ is given by the expression (further we limit the consideration to the case $k \geqslant l$)

$$\langle \delta m_k^\alpha \delta m_l^\beta\rangle = \sum_t \delta m_k^\alpha \delta m_l^\beta w_t, \tag{C1}$$

where the summation is over all possible WL trajectories. Let us consider the trajectory of the length n. The probability $w_t^{(n)}$ of this trajectory is given by the product

$$w_t^{(n)} = \prod_{j=1}^{n} w_{m_j}^{1\to 2} w_{m_j}^{2\to 1} = w_{m_n}^{1\to 2} w_{m_n}^{2\to 1} w_t^{(n-1)}. \tag{C2}$$

For all $n > k$ one can find the recurrent relation

$$\sum_t \delta m_k^\alpha \delta m_l^\beta w_t^{(n)} = \sum_t \delta m_k^\alpha \delta m_l^\beta w_{m_n}^{1\to 2} w_{m_n}^{2\to 1} w_t^{(n-1)} = \sum_t \delta m_k^\alpha \delta m_l^\beta w_t^{(n-1)}, \quad n > k, \tag{C3}$$

where the subscript (n) denotes that the sum is taken over all the possible trajectories of length n. Substituting Eq. (C3) into Eq. (C1) gives

$$\langle \delta m_k^\alpha \delta m_l^\beta\rangle = \delta_{\alpha\beta}\delta_{kl}\sum_t (\delta m_k^\alpha)^2 w_t^{(k)} = \delta_{\alpha\beta}\delta_{kl}\langle (\delta m_k^\alpha)^2\rangle. \tag{C4}$$

The average value $\langle (\delta m_k^{2\to 1})^2\rangle$ is statistically independent of the trajectory and is given by the dispersion of the geometric distribution:

$$\langle (\delta m_k^{2\to 1})^2\rangle = p\sum_{m=1}^{\infty} (m - \langle m\rangle)^2 (1-p)^{m-1} = \frac{q}{p^2}. \tag{C5}$$

The average value $\langle (\delta m_k^{1\to 2})^2\rangle$ depends on the trajectory. However, in the limit $\varepsilon_i \to 0$ the zeroth-order approximation, as follows from Eq. (12), is also given by the Eq. (C5). Using Eq. (C4) one can find $\langle \Delta_\infty\rangle^2$:

$$\langle \Delta_\infty \rangle^2 = \left\langle \left[\varepsilon_i \sum_{l=1}^{\infty} (\delta m_l^{2\to 1} - \delta m_l^{1\to 2})(1-\varepsilon_i/p)^{n-l} \right]^2 \right\rangle = \varepsilon_i^2 \left\langle \sum_{l=1}^{\infty} [(\delta m_l^{2\to 1})^2 + (\delta m_l^{1\to 2})^2](1-\varepsilon_i/p)^{2(n-l)} \right\rangle$$
$$= \varepsilon_i^2 \left\langle \frac{(\delta m_l^{2\to 1})^2 + (\delta m_l^{1\to 2})^2}{1-(1-\varepsilon_i/p)^2} \right\rangle \cong \frac{\varepsilon_i^2}{1-(1-2\varepsilon_i/p)} 2\sigma^2(\delta m_l^{2\to 1}) = \varepsilon_i \frac{q}{p}. \quad \text{(C6)}$$

[1] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[2] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).
[3] B. J. Schulz, K. Binder, M. Müller, and D. P. Landau, Phys. Rev. E **67**, 067102 (2003).
[4] M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos, Phys. Rev. E **66**, 056703 (2002).
[5] B. J. Schulz, K. Binder, and M. Müller, Int. J. Mod. Phys. C **13**, 477 (2002).
[6] N. Rathore, T. A. Knotts IV, and J. J. de Pablo, J. Chem. Phys. **118**, 4285 (2003).
[7] R. G. Ghulghazaryan, Sh. Hayryan, and C. K. Hu, J. Comput. Chem. **28**, 715 (2007).
[8] D. P. Landau, Shan-Ho Tsai, and M. Exler, Am. J. Phys. **72**, 1294 (2004).
[9] C. Zhou and R. N. Bhatt, Phys. Rev. E **72**, 025701(R) (2005).
[10] P. Dayal, S. Trebst, S. Wessel, D. Wurtz, M. Troyer, S. Sabhapandit, and S. N. Coppersmith, Phys. Rev. Lett. **92**, 097201 (2004).
[11] M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos, J. Phys. Chem. B **108**, 19748 (2004).
[12] P. D. Beale, Phys. Rev. Lett. **76**, 78 (1996).